BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 27, 2003, 18:06:39 ; Search time 76 Seconds
(without alignments)
489.170 Million cell updates/sec

Title: US-10-007-267A-8
Perfect score: 1461
Sequence: 1 MQNHVISLASAAERRAHIAA.....REKRRQRREQLIGKIIVPFQ 279

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1461	100.0	279	17	AAR91312	N. gonorrhoeae gly
2	1288	88.2	275	20	AAY22156	N. meningitidis Be
3	1036.5	70.9	280	17	AAR91315	N. gonorrhoeae gly
4	1031.5	70.6	280	18	AAW06580	Lipo-oligosacchari
5	219	15.0	50	20	AAW89328	Neisseria meningit
6	217	14.9	50	20	AAW89329	Neisseria meningit
7	129.5	8.9	423	23	ABB89453	Human polypeptide
8	129.5	8.9	517	21	AAB41790	Human ORFX ORF1554
9	129.5	8.9	517	22	AAM93202	Human polypeptide,
10	129.5	8.9	517	22	AAM93206	Human polypeptide,
11	129.5	8.9	578	22	AAG81257	Human AFP protein
12	129.5	8.9	579	20	AAY30812	Human secreted pro
13	129.5	8.9	595	21	AAB12121	Hydrophobic domain
14	129.5	8.9	636	23	ABP42016	Human ovarian anti
15	121	8.3	1128	22	ABB62795	Drosophila melanog
16	119.5	8.2	521	21	AAB41900	Human ORFX ORF1664
17	119.5	8.2	622	22	AAM93427	Human polypeptide,
18	119.5	8.2	622	22	AAB88452	Human membrane or
19	119.5	8.2	622	23	ABP43477	Human secreted pro
20	106	7.3	201	18	AAW20604	H. pylori cytoplas
21	105.5	7.2	168	22	ABG22090	Novel human diagno
22	99.5	6.8	273	22	AAB60317	Helicobacter pylor
23	98.5	6.7	273	20	AAY32543	H. pylori beta-1,4
24	97.5	6.7	273	22	AAB60321	Helicobacter pylor
25	96.5	6.6	578	21	AAY97423	Rice 1-deoxy-D-xyl
26	95.5	6.5	161	23	AAE15943	Worm C38H2-2/141-3
27	95.5	6.5	389	22	AAB82459	Caenorhabditis ele
28	94	6.4	739	22	ABB11417	Human secreted pro
29	93.5	6.4	444	23	AAU76669	Helicobacter pylor
30	90	6.2	261	22	ABG03682	Novel human diagno
31	89.5	6.1	1378	23	ABB57176	Mouse ischaemic co
32	88.5	6.1	159	21	AAB41784	Human ORFX ORF1548
33	88.5	6.1	500	22	AAU27796	Human full-length
34	88	6.0	651	20	AAY35171	Protein involved i
35	87	6.0	279	18	AAW21023	H. pylori cytoplas
36	86.5	5.9	852	22	ABG08255	Novel human diagno
37	86.5	5.9	955	14	AAR42088	Human p50 or KBF1
38	86.5	5.9	955	14	AAR42234	Human p50 protein
39	86.5	5.9	955	14	AAR42235	Human p50 protein
40	86.5	5.9	955	14	AAR42236	Human p50 protein
41	86.5	5.9	955	14	AAR42237	Human p50 protein
42	86.5	5.9	955	14	AAR42238	Human p50 protein
43	86.5	5.9	955	14	AAR42239	Human p50 protein
44	86.5	5.9	955	14	AAR42240	Human p50 protein
45	86.5	5.9	955	14	AAR42241	Human p50 protein

ALIGNMENTS

RESULT 1
AAR91312
ID AAR91312 standard; Protein; 279 AA.
XX
AC AAR91312;
XX
DT 09-JUL-1996 (first entry)
XX
DE N. gonorrhoeae glycosyltransferase LgtB.
XX
KW Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
KW vaccine.
XX
OS Neisseria gonorrhoeae strain F62.
XX
PN WO9610086-A1.
XX
PD 04-APR-1996.
XX
PF 25-SEP-1995; 95WO-US12317.
XX
PR 26-SEP-1994; 94US-0312387.
XX
PA (UYRQ) UNIV ROCKEFELLER.
XX
PI Gotschlich EC;
XX
DR WPI; 1996-200924/20.
DR P-PSDB; AAT14061.
XX
PT Nucleic acids encoding glycosyl transferase(s) - used in the
PT diagnosis of infection with Neisseria and for the biosynthesis of
PT oligo:saccharide(s)

```
XX
PS   Claim 9; Fig 2c; 81pp; English.
XX
CC   5 Glycosyltransferases (AAR91311-15) are products of the lgt locus
CC   (AAT14061) of Neiserria gonorrhoeae strain F62.  Glycosyltransferase
CC   LgtB (AAR91312) can be obtd. by expression of the lgtB coding
CC   sequence in recombinant host cells.  A method for adding Gal
CC   beta1-4 to GlcNAc or Glc comprises contacting a reaction mixture
CC   contg. activated Gal to an acceptor moiety comprising a GlcNAc
CC   or Glc residue in the presence of LgtB.  Oligosaccharides can be
CC   produced that, when attached to non-toxic lipids, are useful for
CC   Neisseria vaccine prepn.  Blood group core oligosaccharides, and
CC   mimics of lacto-N-neotetraose, gangliosides and saccharide
CC   portions of globoglycolipids can also be produced using the enzymes.
XX
SQ   Sequence   279 AA;

  Query Match             100.0%;  Score 1461;  DB 17;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-159;
  Matches  279;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240

Qy     241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
           |||||||||||||||||||||||||||||||||||||||
Db     241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279


RESULT 2
AAY22156
ID   AAY22156 standard; Protein; 275 AA.
XX
AC   AAY22156;
XX
DT   08-SEP-1999  (first entry)
XX
DE   N. meningitidis Beta-1,4-galactosyltransferase.
XX
KW   Beta-1,4-galactosyltransferase; lgtB; fusion protein; catalytic domain;
KW   glycosyltransferase; accessory enzyme; nucleotide sugar formation;
KW   saccharide donor; oligosaccharide synthesis;
KW   carbohydrate structure development.
XX
OS   Neisseria meningitidis.
XX
PN   WO9931224-A2.
XX
PD   24-JUN-1999.
XX
PF   15-DEC-1998;   98WO-CA01180.
XX
PR   14-DEC-1998;   98US-0211691.
PR   15-DEC-1997;   97US-0069443.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Gilbert M,  Wakarchuk WW,  Young NM;
XX
DR   WPI; 1999-395174/33.
DR   N-PSDB; AAX84281.
XX
PT   A new glycosyltransferase fusion protein useful in the enzymatic
PT   synthesis of oligosaccharides
XX
PS   Example 2; Fig 2; 63pp; English.
XX
CC   This sequence represents the Neisseria meningitidis Beta-1,4-
CC   galactosyltransferase (also referred to as lgtB). The invention relates
CC   to a nucleic acid encoding a fusion protein that comprises a
CC   glycosyltransferase catalytic domain and a catalytic domain from an
CC   accessory enzyme that is involved in formation of a nucleotide sugar
CC   which is a saccharide donor for a glycosyltransferase reaction. The
CC   fusion protein is useful in the enzymatic synthesis of oligosaccharides.
CC   The fusion proteins are able to catalyse more than one reaction involved
CC   in the enzymatic synthesis. This is useful for the development of
CC   therapeutic agents that have specific carbohydrate structures.
CC   Carbohydrates are involved in recognition elements on the surface of
CC   cells. The fusion protein can be used for the synthesis of both natural
CC   carbohydrates and synthetic derivatives with novel properties. The fusion
CC   polypeptide allows two glycosyltransferase reactions in a single vessel,
CC   provides improved yields of end products. Additionally, cleanup and
CC   disposal of extra solvents and by-products is reduced. The fusion protein
CC   can also use directly different donor analogues and various acceptors
CC   with a terminal galactose residue.
XX
SQ   Sequence   275 AA;

  Query Match             88.2%;  Score 1288;  DB 20;  Length 275;
  Best Local Similarity   90.0%;  Pred. No. 3.4e-139;
  Matches  251;  Conservative    7;  Mismatches   17;  Indels    4;  Gaps    1;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||| |||  ||||||||||||||||:||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIADTFGRHGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           |||||||||||||:||||||:||| |||||||||| |:||||| |||||||||:||:||
Db      61 VEKACFMSHAVLWKQALDEGLPYITVFEDDVLLGEGEEKFLAEDAWLQERFDPDTAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||||||||||||||| |||||||||| |||||||||||||||||||| |||| |||
Db     121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMRFFLDRFAALPPEGLHP 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           |||||| :  ||||||||||||||||||||||||||||||||||||||| |||||| |||
Db     181 VDLMMFSDFFDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRLLNRKQQRRDS 240

Qy     241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
           ||||||||||||||| ||||||||||     |||||
Db     241 PANTFKHRLIRALTKISREREKRRQRREQ----FIVPFQ 275


RESULT 3
AAR91315
ID   AAR91315 standard; Protein; 280 AA.
XX
AC   AAR91315;
XX
DT   09-JUL-1996  (first entry)
XX
DE   N. gonorrhoeae glycosyltransferase LgtE.
XX
KW   Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
KW   vaccine.
XX
OS   Neisseria gonorrhoeae strain F62.
XX
PN   WO9610086-A1.
XX
PD   04-APR-1996.
XX
PF   25-SEP-1995;   95WO-US12317.
```

XX
PR 26-SEP-1994; 94US-0312387.
XX
PA (UYRQ) UNIV ROCKEFELLER.
XX
PI Gotschlich EC;
XX
DR WPI; 1996-200924/20.
DR N-PSDB; AAT14061.
XX
PT Nucleic acids encoding glycosyl transferase(s) - used in the
PT diagnosis of infection with Neisseria and for the biosynthesis of
PT oligo:saccharide(s)
XX
PS Claim 12; Fig 2f; 81pp; English.
XX
CC 5 Glycosyltransferases (AAR91311-15) are products of the lgt locus
CC (AAT14061) of Neiserria gonorrhoeae strain F62. Glycosyltransferase
CC LgtE (AAR91315) can be obtd. by expression of the lgtE coding
CC sequence in recombinant host cells. A method for adding Gal
CC beta1-4 to GlcNac or Glc comprises contacting a reaction mixture
CC contg. activated Gal to an acceptor moiety comprising a GlcNAc or
CC Glc residue in the presence of LgtE. Oligosaccharides can be produced
CC that, when attached to non-toxic lipids, are useful for Neisseria
CC vaccine prepn. Blood group core oligosaccharides, and mimics of
CC lacto-N-neotetraose, gangliosides and saccharide portions of
CC globoglycolipids can also be produced using the enzymes.
XX
SQ Sequence 280 AA;

Query Match 70.9%; Score 1036.5; DB 17; Length 280;
Best Local Similarity 71.6%; Pred. No. 2.8e-110;
Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;

```
Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||| ||||||||||||||||||||:|||||||||||||||||
Db       1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||:|||||||||||||||||||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:||||||||||||||||||:|||||||||||||||:
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||    |:||||| |::|||| ||||||||  ||| ||| :| ||   |
Db     181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:: |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280
```

RESULT 4
AAW06580
ID AAW06580 standard; Protein; 280 AA.
XX
AC AAW06580;
XX
DT 21-MAR-1997 (first entry)
XX
DE Lipo-oligosaccharide gene-encoded protein.
XX
KW Polyglycosyltransferase; N-acetylglucosaminyl transferase;
KW N-acetylgalactosaminyl transferase; lipo-oligosaccharide.
XX
OS Neisseria gonorrhoeae ATCC 33084.
XX
PN WO9640971-A1.
XX
PD 19-DEC-1996.
XX
PF 03-JUN-1996; 96WO-US08323.
XX
PR 07-JUN-1995; 95US-0478140.
XX
PA (NEOS-) NEOSE TECHNOLOGIES INC.
XX
PI Buczala SL, Johnson KF, Roth S;
XX
DR WPI; 1997-052351/05.
DR N-PSDB; AAT49230.
XX
PT Transfer of at least 2 saccharide units using
PT poly:glycosyl:transferase - isolated from N. gonorrhoeae, catalyses
PT the addition of both GlcNac and GalNac di:saccharide(s) units to a
PT single galactose moiety
XX
PS Disclosure; Fig 2F-H; 38pp; English.
XX
CC A lipooligosaccharide-encoding gene region (AAT49230) of Neisseria
CC gonorrhoeae ATCC 33084 includes coding sequences for 5 proteins
CC (AAW06576-80), one of which (AAW06576) is a polyglycosyltransferase
CC that catalyses the addition of GlcNAc and GalNAc disaccharides to
CC a galactose moiety. The function of the other 4 proteins is not
CC stated in the specification.
XX
SQ Sequence 280 AA;

Query Match 70.6%; Score 1031.5; DB 18; Length 280;
Best Local Similarity 71.6%; Pred. No. 1.1e-109;
Matches 209; Conservative 23; Mismatches 35; Indels 25; Gaps 3;

```
Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||| ||||||||||||||||||||:|||||||||||||||||
Db       1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||:|||||||||||||||||||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:||||||||||||||||||:|||||||||||||||:
Db     121 LETMFAKVIVRPDFVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||    | ||||| |::|||| ||||||||  ||| ||| :| ||   |
Db     181 VDLMMFTYFFDLEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:: |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280
```

RESULT 5
AAW89328
ID AAW89328 standard; peptide; 50 AA.
XX
AC AAW89328;
XX
DT 26-FEB-1999 (first entry)
XX
DE Neisseria meningitidis lgtB C-terminal peptide.
XX
KW Neisseria meningitidis; lgtC; lgtB; beta-1,4-galactosyltransferase;
KW glycosyltransferase; proteolytic enzyme.
XX
OS Neisseria meningitidis.
XX
PN WO9854331-A2.
XX
PD 03-DEC-1998.
XX
PF 26-MAY-1998; 98WO-IB00975.

```
XX
PR   27-MAY-1997;   97US-0047751.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Wakarchuk WW,  Young NM;
XX
DR   WPI; 1999-035177/03.
XX
PT   Expressing high levels of glycosyltransferases - comprises use of
PT   either host cells deficient in proteolytic enzymes or modified
PT   glycosyltransferase genes deleted in a proteolytic recognition site
XX
PS   Example 1; Fig 8; 61pp; English.
XX
CC   A method has been developed of expressing a glycosyltransferase in a
CC   host cell. The method comprises introducing into the host cell a nucleic
CC   acid encoding the glycosyltransferase and incubating the host cell under
CC   conditions appropriate for expression of the glycosyltransferase, where
CC   the host cell substantially lacks a protease that cleaves polypeptides
CC   between two consecutive positively charged amino acid residues. The
CC   glycosyltransferase can be used in in vitro production of
CC   oligosaccharide structures which are potential therapeutic agents for
CC   use in the manipulation of cell-cell recognition events, particularly
CC   adhesion of bacteria and viruses to mammalian cells and leukocyte-
CC   endothelial cell interaction through selectins in inflammation. The
CC   method provides more readily recoverable active glycosyltransferases
CC   than prior art methods involving mammalian glycosyltransferases. The
CC   present sequence represents a C-terminal peptide from Neisseria
CC   meningitidis lgtB from the present invention.
XX
SQ   Sequence   50 AA;

  Query Match             15.0%;  Score 219;  DB 20;  Length 50;
  Best Local Similarity   85.2%;  Pred. No. 2e-17;
  Matches   46;  Conservative    0;  Mismatches    4;  Indels    4;  Gaps    1;

Qy    226 EHDRRLNRKQQWRDSPANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          |||| |||||| |||||||||||||||||| ||||||||||||     |||||
Db      1 EHDRLLNRKQQRRDSPANTFKHRLIRALTKISREREKRRQRREQ----FIVPFQ 50


RESULT 6
AAW89329
ID   AAW89329 standard; peptide; 50 AA.
XX
AC   AAW89329;
XX
DT   26-FEB-1999  (first entry)
XX
DE   Neisseria meningitidis lgtB C-terminal peptide.
XX
KW   Neisseria meningitidis; lgtC; lgtB; beta-1,4-galactosyltransferase;
KW   glycosyltransferase; proteolytic enzyme.
XX
OS   Neisseria meningitidis.
XX
PN   WO9854331-A2.
XX
PD   03-DEC-1998.
XX
PF   26-MAY-1998;   98WO-IB00975.
XX
PR   27-MAY-1997;   97US-0047751.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Wakarchuk WW,  Young NM;
XX
DR   WPI; 1999-035177/03.
XX
PT   Expressing high levels of glycosyltransferases - comprises use of
PT   either host cells deficient in proteolytic enzymes or modified
PT   glycosyltransferase genes deleted in a proteolytic recognition site
XX
PS   Example 1; Fig 8; 61pp; English.
XX
CC   A method has been developed of expressing a glycosyltransferase in a
CC   host cell. The method comprises introducing into the host cell a nucleic
CC   acid encoding the glycosyltransferase and incubating the host cell under
CC   conditions appropriate for expression of the glycosyltransferase, where
CC   the host cell substantially lacks a protease that cleaves polypeptides
CC   between two consecutive positively charged amino acid residues. The
CC   glycosyltransferase can be used in in vitro production of
CC   oligosaccharide structures which are potential therapeutic agents for
CC   use in the manipulation of cell-cell recognition events, particularly
CC   adhesion of bacteria and viruses to mammalian cells and leukocyte-
CC   endothelial cell interaction through selectins in inflammation. The
CC   method provides more readily recoverable active glycosyltransferases
CC   than prior art methods involving mammalian glycosyltransferases. The
CC   present sequence represents a C-terminal peptide from Neisseria
CC   meningitidis lgtB from the present invention.
XX
SQ   Sequence   50 AA;

  Query Match             14.9%;  Score 217;  DB 20;  Length 50;
  Best Local Similarity   83.3%;  Pred. No. 3.4e-17;
  Matches   45;  Conservative    0;  Mismatches    5;  Indels    4;  Gaps    1;

Qy    226 EHDRRLNRKQQWRDSPANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          |||| |||||||||||||||| ||||||||| |||| |||||||     |||||
Db      1 EHDRLLNRKQQWRDSPANTFKRRLIRALTKISREREIRRQRREQ----FIVPFQ 50


RESULT 7
ABB89453
ID   ABB89453 standard; Protein; 423 AA.
XX
AC   ABB89453;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polypeptide SEQ ID NO 1829.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-US16450.
XX
PR   19-MAY-2000; 2000US-205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   N-PSDB; ABL89862.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders -
XX
PS   Claim 11; SEQ ID NO 1829; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
```

```
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   423 AA;

  Query Match             8.9%;  Score 129.5;  DB 23;  Length 423;
  Best Local Similarity   27.2%;  Pred. No. 9.3e-06;
  Matches   65;  Conservative   30;  Mismatches   91;  Indels   53;  Gaps   12;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQFFDA----LMPSERLERAMAELVPGLSAHPY--- 57
           |||||   :||  : |:     |  :  ||    :: |  :     :|:||    ||
Db     155 VISLARRPDRRERMLASLWEMEISGRVVDAVDGWMLNSSAIRNLGVDLLPGYQ-DPYSGR 213

Qy      58 -LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSA 116
            |:  |  ||:|| :||: :  |:  : ||||||                 ||  :|
Db     214 TLTKGEVGCFLSHYSIWEEVVARGLARVLVFEDDV-----------------RF--ESN 253

Qy     117 FVVRLETMFMHVLTSP-SGVADYGGR--AFPLLESEHCGTAGYIIS-------RKAMRFF 166
           |  ||| :   |     |    | ||    |  |:   |  | :::         |:|
Db     254 FRGRLERLMEDVEAEKLSWDLIYLGRKQVNPEKETAVEGLPGLVVAGYSYWTLAYALRLA 313

Qy     167 LDR--FAVLPPERLHPVD----LMMFGNPDDR-------EGMPVCQLNPALCAQELHYA 212
           |   |  |  |: |||    :|   :|:::         :      | | |   |||
Db     314 GARKLLASQPLRRMLPVDEFLPIMFDQHPNEQYKAHFWPRDLVAFSAQPLLAA-PTHYA 371


RESULT 8
AAB41790
ID   AAB41790 standard; Protein; 517 AA.
XX
AC   AAB41790;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF1554 polypeptide sequence SEQ ID NO:3108.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; AAC75999.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 2331-2332; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence   517 AA;

  Query Match             8.9%;  Score 129.5;  DB 21;  Length 517;
  Best Local Similarity   27.2%;  Pred. No. 1.3e-05;
  Matches   65;  Conservative   30;  Mismatches   91;  Indels   53;  Gaps   12;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQFFDA----LMPSERLERAMAELVPGLSAHPY--- 57
           |||||   :||  : |:     |  :  ||    :: |  :     :|:||    ||
Db     246 VISLARRPDRRERMLASLWEMEISGRVVDAVDGWMLNSSAIRNLGVDLLPGYQ-DPYSGR 304

Qy      58 -LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSA 116
            |:  |  ||:||  :||: :  |:  : ||||||                 ||  :|
Db     305 TLTKGEVGCFLSHYSIWEEVVARGLARVLVFEDDV-----------------RF--ESN 344

Qy     117 FVVRLETMFMHVLTSP-SGVADYGGR--AFPLLESEHCGTAGYIIS-------RKAMRFF 166
           |  ||| :   |     |    | ||    |  |:   |  | :::         |:|
Db     345 FRGRLERLMEDVEAEKLSWDLIYLGRKQVNPEKETAVEGLPGLVVAGYSYWTLAYALRLA 404

Qy     167 LDR--FAVLPPERLHPVD----LMMFGNPDDR-------EGMPVCQLNPALCAQELHYA 212
           |   |  |  |: |||    :|   :|:::         :      | | |   |||
Db     405 GARKLLASQPLRRMLPVDEFLPIMFDQHPNEQYKAHFWPRDLVAFSAQPLLAA-PTHYA 462


RESULT 9
AAM93202
ID   AAM93202 standard; Protein; 517 AA.
XX
AC   AAM93202;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 2591.
XX
```

```
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-0114089.
XX
PR   08-JUL-1999;   99JP-0194486.
PR   11-JAN-2000; 2000JP-0118774.
PR   02-MAY-2000; 2000JP-0183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94110.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their
PT   use in genetic manipulation -
XX
PS   Claim 8; SEQ ID NO 2591; 1380pp + sequence listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been
CC   isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC   molecules have been determined. Primers for synthesising the full length
CC   cDNA are useful for clarifying the function of the protein encoded by
CC   the cDNA. The full length clones were obtained by construction of full
CC   length enriched cDNA libraries that were synthesised by the oligo-capping
CC   method. The primers enable the production of the full length cDNA easily
CC   without any special methods. The present sequence is a polypeptide
CC   encoded by a full length human cDNA of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ   Sequence   517 AA;

  Query Match             8.9%;  Score 129.5;  DB 22;  Length 517;
  Best Local Similarity   27.2%;  Pred. No. 1.3e-05;
  Matches   65;  Conservative   30;  Mismatches   91;  Indels   53;  Gaps   12;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQFFDA----LMPSERLERAMAELVPGLSAHPY--- 57
           |||||   :||  : |:     |  :  ||    :: |  :     :|:||    ||
Db     246 VISLARRPDRRERMLASLWEMEISGRVVDAVDGWMLNSSAIRNLGVDLLPGYQ-DPYSGR 304

Qy      58 -LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSA 116
            |:  |  ||:||  :||: :  |:  : ||||||                 ||  :|
Db     305 TLTKGEVGCFLSHYSIWEEVVARGLARVLVFEDDV-----------------RF--ESN 344

Qy     117 FVVRLETMFMHVLTSP-SGVADYGGR--AFPLLESEHCGTAGYIIS-------RKAMRFF 166
           |  ||| :  |     |    | ||    |  |:   |  | :::         |:|
Db     345 FRGRLERLMEDVEAEKLSWDLIYLGRKQVNPEKETAVEGLPGLVVAGYSYWTLAYALRLA 404

Qy     167 LDR--FAVLPPERLHPVD----LMMFGNPDDR-------EGMPVCQLNPALCAQELHYA 212
           |   |  |  |: |||    :|   :|:::         :      | |.|   |||
Db     405 GARKLLASQPLRRMLPVDEFLPIMFDQHPNEQYKAHFWPRDLVAFSAQPLLAA-PTHYA 462


RESULT 10
AAM93206
ID   AAM93206 standard; Protein; 517 AA.
XX
AC   AAM93206;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 2599.
XX
```

```
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-0114089.
XX
PR   08-JUL-1999;   99JP-0194486.
PR   11-JAN-2000; 2000JP-0118774.
PR   02-MAY-2000; 2000JP-0183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94114.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their
PT   use in genetic manipulation -
XX
PS   Claim 8; SEQ ID NO 2599; 1380pp + sequence listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been
CC   isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC   molecules have been determined. Primers for synthesising the full length
CC   cDNA are useful for clarifying the function of the protein encoded by
CC   the cDNA. The full length clones were obtained by construction of full
CC   length enriched cDNA libraries that were synthesised by the oligo-capping
CC   method. The primers enable the production of the full length cDNA easily
CC   without any special methods. The present sequence is a polypeptide
CC   encoded by a full length human cDNA of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ   Sequence   517 AA;

  Query Match             8.9%;  Score 129.5;  DB 22;  Length 517;
  Best Local Similarity   27.2%;  Pred. No. 1.3e-05;
  Matches   65;  Conservative   30;  Mismatches   91;  Indels   53;  Gaps   12;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQFFDA----LMPSERLERAMAELVPGLSAHPY--- 57
           |||||   :||  : |:     |  :  ||    :: |  :     :|:||    ||
Db     246 VISLARRPDRRERMLASLWEMEISGRVVDAVDGWMLNSSAIRNLGVDLLPGYQ-DPYSGR 304

Qy      58 -LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSA 116
            |:  |  ||:||  :||: :  |:  : ||||||                 ||  :|
Db     305 TLTKGEVGCFLSHYSIWEEVVARGLARVLVFEDDV-----------------RF--ESN 344

Qy     117 FVVRLETMFMHVLTSP-SGVADYGGR--AFPLLESEHCGTAGYIIS-------RKAMRFF 166
           |  ||| :  |     |    | ||    |  |:   |  | :::         |:|
Db     345 FRGRLERLMEDVEAEKLSWDLIYLGRKQVNPEKETAVEGLPGLVVAGYSYWTLAYALRLA 404

Qy     167 LDR--FAVLPPERLHPVD----LMMFGNPDDR-------EGMPVCQLNPALCAQELHYA 212
           |   |  |  |: |||    :|   :|:::         :      | |.|   |||
Db     405 GARKLLASQPLRRMLPVDEFLPIMFDQHPNEQYKAHFWPRDLVAFSAQPLLAA-PTHYA 462


RESULT 11
AAG81257
ID   AAG81257 standard; Protein; 578 AA.
XX
AC   AAG81257;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human AFP protein sequence SEQ ID NO:32.
XX
```

```
KW   Human; secreted protein; secretion; bacterial cell; fungal cell;
KW   eukaryotic cell; fusion protein; maltose binding protein;
KW   immunoglobulin constant region; polyhistidine tag.
XX
OS   Homo sapiens.
XX
PN   WO200129221-A2.
XX
PD   26-APR-2001.
XX
PF   20-OCT-2000; 2000WO-US29052.
XX
PR   20-OCT-1999;   99US-0160712.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC,  Yee DP;
XX
DR   WPI; 2001-300340/31.
DR   N-PSDB; AAH52108.
XX
PT   Isolated polypeptide for directing secretion of proteins of interest
PT   from a host cell including, e.g. bacteria, includes contiguous amino
PT   acid residues of polypeptide with specified amino acids  -
XX
PS   Claim 1; Page 121-122; 617pp; English.
XX
CC   AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242
CC   to AAG81453. The secreted proteins can be used for directing the
CC   secretion of proteins of interest from a host cell including bacteria,
CC   fungal cells, and cultured higher eukaryotic cells. The present invention
CC   also describes fusion proteins, where a secreted protein of the invention
CC   is operably linked via a peptide bond or peptide linker to a second
CC   protein selected from the group consisting of maltose binding protein,
CC   an immunoglobulin constant region, a polyhistidine tag and a peptide
CC   given in AAG81453.
XX
SQ   Sequence   578 AA;

  Query Match             8.9%;  Score 129.5;  DB 22;  Length 578;
  Best Local Similarity   27.2%;  Pred. No. 1.5e-05;
  Matches   65;  Conservative   30;  Mismatches   91;  Indels   53;  Gaps   12;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQFFDA----LMPSERLERAMAELVPGLSAHPY--- 57
           |||||   :||  : |:     |  :  ||    :: |  :     :|:||    ||
Db     324 VISLARRPDRRERMLASLWEMEISGRVVDAVDGWMLNSSAIRNLGVDLLPGYQ-DPYSGR 382

Qy      58 -LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSA 116
            |:  |  ||:||  :||: :  |:  : ||||||                  ||  :|
Db     383 TLTKGEVGCFLSHYSIWEEVVARGLARVLVFEDDV-----------------RF--ESN 422

Qy     117 FVVRLETMFMHVLTSP-SGVADYGGR--AFPLLESEHCGTAGYIIS-------RKAMRFF 166
           |  ||| :   |     |    | ||    |  |:   |  | :::         |:|
Db     423 FRGRLERLMEDVEAEKLSWDLIYLGRKQVNPEKETAVEGLPGLVVAGYSYWTLAYALRLA 482

Qy     167 LDR--FAVLPPERLHPVD----LMMFGNPDDR-------EGMPVCQLNPALCAQELHYA 212
             |   |  |  |: |||    :|   :|:::        :      | | |   |||
Db     483 GARKLLASQPLRRMLPVDEFLPIMFDQHPNEQYKAHFWPRDLVAFSAQPLLAA-PTHYA 540


RESULT 12
AAY30812
ID   AAY30812 standard; Protein; 579 AA.
XX
AC   AAY30812;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human secreted protein encoded from gene 2.
XX
KW   Secreted protein; prevention; treatment; protein therapy; gene therapy;
KW   diagnosis; cancer; tumour; neurodegenerative disorder; blood disorder;
KW   developmental abnormality; fetal deficiency; leukemia; autoimmune; acne;
KW   hepatic disease; renal disease; lymphoma; inflammation; allergy; asthma;
KW   Alzheimer's disease; cognitive disorder; schizophrenia; obesity; sepsis;
KW   osteoporosis; arthritis; infection; AIDS; connective tissue disorder;
KW   transplant rejection; diabetes; psoriasis; cardiovascular disorder;
KW   reproductive disorder; food additive; food preservative; human; primer;
KW   early promoter; GAS; gamma activation element.
XX
OS   Homo sapiens.
XX
PN   WO9940100-A1.
XX
PD   12-AUG-1999.
XX
PF   04-FEB-1999;   99WO-US02293.
XX
PR   09-FEB-1998;   98US-0074341.
PR   09-FEB-1998;   98US-0074037.
PR   09-FEB-1998;   98US-0074118.
PR   09-FEB-1998;   98US-0074141.
PR   09-FEB-1998;   98US-0074157.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Kyaw H,  Lafleur DW,  Moore PA,  Rosen CA,  Ruben SM;
PI   Shi Y,  Wei Y;
XX
DR   WPI; 1999-479426/40.
DR   N-PSDB; AAZ00803.
XX
PT   New isolated human genes potentially useful for, e.g. developmental
PT   abnormalities and fetal deficiencies
XX
PS   Claim 1b; Page 206-208; 263pp; English.
XX
CC   This invention describes novel isolated human genes and the secreted
CC   proteins they encode. The polynucleotides and their corresponding
CC   secreted polypeptides are useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. Also pathological
CC   conditions can be diagnosed by determining the amount of the new
CC   polypeptides in a sample or by determining the presence of mutations in
CC   the new polynucleotides. Specific uses are described for the
CC   polynucleotides of the invention based on which tissues they are most
CC   highly expressed in, and include developing products for the diagnosis or
CC   treatment of cancer, tumours, neurodegenerative disorders, developmental
CC   abnormalities and fetal deficiencies, blood disorders, leukemias,
CC   diseases of the immune system, autoimmune diseases, hepatic and renal
CC   disease, lymphomas, inflammation, allergies, Alzheimer's and cognitive
CC   disorders, schizophrenia, obesity, osteoporosis, arthritis, infections,
CC   AIDS, connective tissue disorders, transplant rejection, diabetes,
CC   asthma, sepsis, acne, psoriasis, cardiovascular disorders, and
CC   reproductive disorders. The polypeptides or polynucleotides can also be
CC   used as food additives or preservatives. The polypeptide are also useful
CC   for identifying their binding partners. This sequence represents a
CC   secreted protein described in the invention.
XX
SQ   Sequence   579 AA;

  Query Match             8.9%;  Score 129.5;  DB 20;  Length 579;
  Best Local Similarity   27.2%;  Pred. No. 1.5e-05;
  Matches   65;  Conservative   30;  Mismatches   91;  Indels   53;  Gaps   12;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQFFDA----LMPSERLERAMAELVPGLSAHPY--- 57
           |||||   :||  : |:     |  :  ||    :: |  :     :|:||    ||
Db     324 VISLARRPDRRERMLASLWEMEISGRVVDAVDGWMLNSSAIRNLGVDLLPGYQ-DPYSGR 382

Qy      58 -LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSA 116
            |:  |  ||:||  :||: :  |:  : ||||||                  ||  :|
Db     383 TLTKGEVGCFLSHYSIWEEVVARGLARVLVFEDDV-----------------RF--ESN 422

Qy     117 FVVRLETMFMHVLTSP-SGVADYGGR--AFPLLESEHCGTAGYIIS-------RKAMRFF 166
           |  ||| :   |     |    | ||    |  |:   |  | :::         |:|
Db     423 FRGRLERLMEDVEAEKLSWDLIYLGRKQVNPEKETAVEGLPGLVVAGYSYWTLAYALRLA 482
```

```
Qy    167 LDR--FAVLPPERLHPVD----LMMFGNPDDR-------EGMPVCQLNPALCAQELHYA 212
          |   | | |: |||    :|   :|:::       :      | | |   |||
Db    483 GARKLLASQPLRRMLPVDEFLPIMFDQHPNEQYKAHFWPRDLVAFSAQPLLAA-PTHYA 540


RESULT 13
AAB12121
ID   AAB12121 standard; Protein; 595 AA.
XX
AC   AAB12121;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Hydrophobic domain protein from clone HP02962 isolated from KB cells.
XX
KW   Human; secreted protein; membrane protein; hydrophobic domain;
KW   proliferation control; differentiation induction; material transport;
KW   biophylaxis; signal receptor; ion channel; transporter; immunostimulant;
KW   immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic;
KW   haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW   autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200029448-A2.
XX
PD   25-MAY-2000.
XX
PF   17-NOV-1999;   99WO-JP06412.
XX
PR   17-NOV-1998;   98JP-0326255.
PR   22-DEC-1998;   98JP-0364315.
PR   16-MAR-1999;   99JP-0069811.
PR   27-APR-1999;   99JP-0119299.
PR   19-MAY-1999;   99JP-0138169.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2000-387753/33.
DR   N-PSDB; AAA60183, AAA60193.
XX
PT   Proteins comprising hydrophobic regions, such as secretory and membrane
PT   proteins, useful in research and diagnostics and having various
PT   activities e.g. immunomodulatory, antiinflammatory, chemokinetic,
PT   hemostatic, thrombolytic -
XX
PS   Claim 1; Page 184-186; 410pp; English.
XX
CC   Secretory proteins play important roles in the proliferation control, the
CC   differentiation induction, the material transport and the biophylaxis of
CC   cells. Membrane proteins have important roles as signal receptors, ion
CC   channels and transporters. The present sequence is a human protein which
CC   has at least one hydrophobic domain. This protein may be a secretory or a
CC   membrane protein. The present protein may have cytokine and cell
CC   proliferation/differentiation activity, immune stimulating or suppressing
CC   activity, haematopoiesis activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, anti-inflammatory activity and tumour
CC   inhibition activity. The present protein could therefore be used for
CC   treatment of autoimmune disease, Alzheimer's disease, Parkinson's
CC   disease, and cancer.
XX
SQ   Sequence   595 AA;

  Query Match             8.9%;  Score 129.5;  DB 21;  Length 595;
  Best Local Similarity   27.2%;  Pred. No. 1.5e-05;
  Matches   65;  Conservative   30;  Mismatches   91;  Indels   53;  Gaps   12;

Qy      5 VISLASAAERRAHIAATFGSRGIPFQFFDA----LMPSERLERAMAELVPGLSAHPY--- 57
          |||||   :||  : |:     |  :  ||    :: |  :     :|:||    ||

Db    324 VISLARRPDRRERMLASLWEMEISGRVVDAVDGWMLNSSAIRNLGVDLLPGYQ-DPYSGR 382

Qy     58 -LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSA 116
           |:  |  ||:||  :||: :  |:  : ||||||                  ||  :|
Db    383 TLTKGEVGCFLSHYSIWEEVVARGLARVLVFEDDV-----------------RF--ESN 422

Qy    117 FVVRLETMFMHVLTSP-SGVADYGGR--AFPLLESEHCGTAGYIIS-------RKAMRFF 166
          |  ||| :   |     |    | ||    |  |:   |  | :::         |:|
Db    423 FRGRLERLMEDVEAEKLSWDLIYLGRKQVNPEKETAVEGLPGLVVAGYSYWTLAYALRLA 482

Qy    167 LDR--FAVLPPERLHPVD----LMMFGNPDDR-------EGMPVCQLNPALCAQELHYA 212
            |   | | |: |||    :|   :|:::       :      | | |   |||
Db    483 GARKLLASQPLRRMLPVDEFLPIMFDQHPNEQYKAHFWPRDLVAFSAQPLLAA-PTHYA 540


RESULT 14
ABP42016
ID   ABP42016 standard; Protein; 636 AA.
XX
AC   ABP42016;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human ovarian antigen HFKHF51, SEQ ID NO:3148.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; drug screening;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive.
XX
OS   Homo sapiens.
XX
PN   WO200200677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US18569.
XX
PR   07-JUN-2000; 2000US-209467P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-147878/19.
DR   N-PSDB; ABQ55093.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g.
PT   ovarian cancer), immune disorders, cardiovascular disorders and
PT   neurological diseases -
XX
PS   Claim 11; SEQ ID No 3148; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related
CC   disorders. Such conditions include ovarian cancer and breast cancer, and
CC   metastatic tumours of ovarian or breast origin, reproductive system
CC   disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC   polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC   disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
```

```
CC   shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC   vaginitis), immune disorders (e.g., congenital and acquired
CC   immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC   blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC   respiratory disorders, neurological disorders, gastrointestinal disorders
CC   and urinary system disorders. Ovarian antigen polypeptides and
CC   polynucleotides may also be used in screening for compounds which
CC   modulate ovarian antigen expression or activity. The polynucleotides may
CC   further be used for gene therapy, chromosome mapping, in the
CC   identification of individuals and in forensic analysis, and the
CC   polypeptides may be used as food additives or to prepare antibodies
CC   useful in disease diagnosis, drug targeting and phenotyping. The present
CC   sequence represents a human ovarian antigen of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   636 AA;

  Query Match             8.9%;  Score 129.5;  DB 23;  Length 636;
  Best Local Similarity   27.2%;  Pred. No. 1.7e-05;
  Matches   65;  Conservative   30;  Mismatches   91;  Indels   53;  Gaps   12;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQFFDA----LMPSERLERAMAELVPGLSAHPY--- 57
           |||||   :||  : |:      |  :  ||    :: |  :     :|:||    ||
Db     365 VISLARRPDRRERMLASLWEMEISGRVVDAVDGWMLNSSAIRNLGVDLLPGYQ-DPYSGR 423

Qy      58 -LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSA 116
            |:  |  ||:||  :||: :  |:  : ||||||                ||  :|
Db     424 TLTKGEVGCFLSHYSIWEEVVARGLARVLVFEDDV-----------------RF--ESN 463

Qy     117 FVVRLETMFMHVLTSP-SGVADYGGR--AFPLLESEHCGTAGYIIS-------RKAMRFF 166
           |  ||| :   |     |    | ||    |  |:   |  |':::         |:|
Db     464 FRGRLERLMEDVEAEKLSWDLIYLGRKQVNPEKETAVEGLPGLVVAGYSYWTLAYALRLA 523

Qy     167 LDR--FAVLPPERLHPVD----LMMFGNPDDR--------EGMPVCQLNPALCAQELHYA 212
             |   |  |  |: |||    :|   :|:::         :      | | |   |||
Db     524 GARKLLASQPLRRMLPVDEFLPIMFDQHPNEQYKAHFWPRDLVAFSAQPLLAA-PTHYA 581


RESULT 15
ABB62795
ID   ABB62795 standard; Protein; 1128 AA.
XX
AC   ABB62795;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 15177.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06898.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 15177; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1128 AA;

  Query Match             8.3%;  Score 121;  DB 22;  Length 1128;
  Best Local Similarity   24.7%;  Pred. No. 0.00038;
  Matches   53;  Conservative   28;  Mismatches   78;  Indels   56;  Gaps    9;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQFFDAL----MPSERLERAMAELVPGLS---AHPY 57
           :|:|    |||  :   |   ||  : | |:    : :|||       :||      |
Db     853 MINLKRRPERREKMERLFDEIGIEAEHFPAVDGKELSTERLLEMGVRFLPGYEDPYHHRA 912

Qy      58 LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAF 117
           ::  |  ||:||  :|   : : :  : : |||:                   ||:|
Db     913 MTMGEIGCFLSHYNIWVMMVRKQLKEVLILEDDI-----------------RFEP---- 950

Qy     118 VVRLETMFMHVLTSPSGVADYG----GRAFPLLESE----------HCG----TAGYIIS 159
             : :|       | |      ||     |||           | |    | ||:||
Db     951 --YFRQNAVRILNQARNAAQYDLIYFGRKRLKEESEPAVENADNLVHAGYSYWTLGYVIS 1008

Qy     160 RKAMRFFLDRFAVLPPERLHPVD----LMMFGNPD 190
             ::  |   |  | ::| |||    ||   :|:
Db    1009 ---LQGALKLLAAKPLDKLIPVDEFLPLMFDRHPN 1040


Search completed: May 27, 2003, 18:15:43
Job time : 78 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 27, 2003, 18:14:10 ; Search time 29 Seconds
(without alignments)
283.069 Million cell updates/sec

Title: US-10-007-267A-8
Perfect score: 1461
Sequence: 1 MQNHVISLASAAERRAHIAA.....REKRRQRREQLIGKIIVPFQ 279

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1461	100.0	279	1	US-08-312-387B-8	Sequence 8, Appli
2	1461	100.0	279	1	US-08-683-426-8	Sequence 8, Appli
3	1461	100.0	279	1	US-08-683-458-8	Sequence 8, Appli
4	1461	100.0	279	2	US-08-878-360-8	Sequence 8, Appli
5	1461	100.0	279	4	US-09-333-412-8	Sequence 8, Appli
6	1036.5	70.9	280	1	US-08-312-387B-6	Sequence 6, Appli
7	1036.5	70.9	280	1	US-08-683-426-6	Sequence 6, Appli
8	1036.5	70.9	280	1	US-08-683-458-6	Sequence 6, Appli
9	1036.5	70.9	280	2	US-08-878-360-6	Sequence 6, Appli
10	1036.5	70.9	280	3	US-08-478-140B-6	Sequence 6, Appli
11	1036.5	70.9	280	4	US-09-333-412-6	Sequence 6, Appli
12	1036.5	70.9	280	4	US-09-338-943-6	Sequence 6, Appli
13	82	5.6	304	2	US-08-701-191A-27	Sequence 27, Appl
14	81.5	5.6	337	4	US-09-134-001C-3799	Sequence 3799, Ap
15	81	5.5	554	4	US-09-564-805-236	Sequence 236, App
16	80.5	5.5	540	3	US-08-964-268-6	Sequence 6, Appli
17	80	5.5	285	4	US-09-186-276B-54	Sequence 54, Appl
18	80	5.5	285	4	US-08-842-445-54	Sequence 54, Appl
19	80	5.5	285	4	US-09-186-188B-54	Sequence 54, Appl
20	79.5	5.4	969	1	US-08-365-689-3	Sequence 3, Appli
21	79.5	5.4	969	1	US-07-747-781-3	Sequence 3, Appli
22	79.5	5.4	969	1	US-08-145-138A-3	Sequence 3, Appli
23	79.5	5.4	969	5	PCT-US92-06888-3	Sequence 3, Appli
24	79.5	5.4	969	5	PCT-US93-03027-4	Sequence 4, Appli
25	78.5	5.4	316	1	US-08-278-089A-16	Sequence 16, Appl
26	78.5	5.4	316	2	US-08-838-957A-15	Sequence 15, Appl
27	78.5	5.4	501	1	US-07-687-466B-2	Sequence 2, Appli
28	78.5	5.4	501	1	US-08-434-702-2	Sequence 2, Appli
29	78.5	5.4	501	1	US-08-271-883-2	Sequence 2, Appli
30	78.5	5.4	501	6	5168064-4	Patent No. 516806
31	75	5.1	628	3	US-08-776-271-2	Sequence 2, Appli
32	75	5.1	628	4	US-09-215-035-2	Sequence 2, Appli
33	74.5	5.1	340	1	US-08-606-888A-5	Sequence 5, Appli
34	73.5	5.0	269	2	US-07-857-224B-79	Sequence 79, Appl
35	73	5.0	261	1	US-08-245-688-6	Sequence 6, Appli
36	73	5.0	586	3	US-08-964-268-3	Sequence 3, Appli
37	73	5.0	738	4	US-08-989-385-1	Sequence 1, Appli
38	72	4.9	261	1	US-08-245-688-2	Sequence 2, Appli
39	72	4.9	261	1	US-08-245-688-12	Sequence 12, Appl
40	71.5	4.9	124	2	US-08-743-200-12	Sequence 12, Appl
41	71.5	4.9	377	2	US-08-969-106-6	Sequence 6, Appli
42	71.5	4.9	377	4	US-09-054-492B-1	Sequence 1, Appli
43	70.5	4.8	256	4	US-09-355-166-2	Sequence 2, Appli
44	70.5	4.8	329	4	US-09-071-035-368	Sequence 368, App
45	70.5	4.8	357	4	US-09-071-035-366	Sequence 366, App

ALIGNMENTS

RESULT 1
US-08-312-387B-8
; Sequence 8, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-312-387B-8

Query Match 100.0%; Score 1461; DB 1; Length 279;
Best Local Similarity 100.0%; Pred. No. 4.1e-160;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240

Qy     241 PANTFKHRLIRALTKIGRREKRRQRREQLIGKIIVPFQ 279
                ||||||||||||||||||||||||||||||||||||||
Db     241 PANTFKHRLIRALTKIGRREKRRQRREQLIGKIIVPFQ 279


RESULT 2
US-08-683-426-8
; Sequence 8, Application US/08683426
; Patent No. 5705367
;  GENERAL INFORMATION:
;    APPLICANT:  Gotschlich, Emil C.
;    TITLE OF INVENTION:  GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;    TITLE OF INVENTION:  OLIGOSACCHARIDES, AND GENES ENCODING THEM
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/683,426
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/312,387
;      FILING DATE:  September 26, 1994
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-095B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  279 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-683-426-8

  Query Match             100.0%;  Score 1461;  DB 1;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-160;
  Matches  279;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240

Qy     241 PANTFKHRLIRALTKIGRREKRRQRREQLIGKIIVPFQ 279
                ||||||||||||||||||||||||||||||||||||||
Db     241 PANTFKHRLIRALTKIGRREKRRQRREQLIGKIIVPFQ 279


RESULT 3
US-08-683-458-8
; Sequence 8, Application US/08683458
; Patent No. 5798233
;  GENERAL INFORMATION:
;    APPLICANT:  Gotschlich, Emil C.
;    TITLE OF INVENTION:  GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;    TITLE OF INVENTION:  OLIGOSACCHARIDES, AND GENES ENCODING THEM
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/683,458
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/312,387
;      FILING DATE:  September 26, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-095A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  279 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-683-458-8

  Query Match             100.0%;  Score 1461;  DB 1;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-160;
  Matches  279;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
```

```
Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240

Qy    241 PANTFKHRLIRALTKIGRERERKRRQRREQLIGKIIVPFQ 279
          |||||||||||||||||||||||||||||||||||||||
Db    241 PANTFKHRLIRALTKIGRERERKRRQRREQLIGKIIVPFQ 279


RESULT 4
US-08-878-360-8
; Sequence 8, Application US/08878360
; Patent No. 5945322
;  GENERAL INFORMATION:
;    APPLICANT:  Gotschlich, Emil C.
;    TITLE OF INVENTION:  GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;    TITLE OF INVENTION:  OLIGOSACCHARIDES, AND GENES ENCODING THEM
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,360
;      FILING DATE:  18-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/683,426
;      FILING DATE:
;      APPLICATION NUMBER:  08/312,387
;      FILING DATE:  September 26, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-095B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  279 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-878-360-8

  Query Match           100.0%;  Score 1461;  DB 2;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-160;
  Matches  279;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240

Qy    241 PANTFKHRLIRALTKIGRERERKRRQRREQLIGKIIVPFQ 279
          |||||||||||||||||||||||||||||||||||||||
Db    241 PANTFKHRLIRALTKIGRERERKRRQRREQLIGKIIVPFQ 279


RESULT 5
US-09-333-412-8
; Sequence 8, Application US/09333412
; Patent No. 6342382
;   GENERAL INFORMATION:
;        APPLICANT: Gotschlich, Emil C.
;        TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;                            OLIGOSACCHARIDES, AND GENES ENCODING THEM
;        NUMBER OF SEQUENCES: 12
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Klauber & Jackson
;             STREET: 411 Hackensack Avenue
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/333,412
;             FILING DATE: 15-Jun-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/312,387
;             FILING DATE: July 7, 1994
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jackson Esq., David A.
;             REGISTRATION NUMBER: 26,742
;             REFERENCE/DOCKET NUMBER: 600-1-095
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201 487-5800
;             TELEFAX: 201 343-1684
;             TELEX: 133521
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 279 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-333-412-8

  Query Match           100.0%;  Score 1461;  DB 4;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-160;
  Matches  279;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
```

```
Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          ||||||||||||||||||||||||||||||||||||||||
Db    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279


RESULT 6
US-08-312-387B-6
; Sequence 6, Application US/08312387B
; Patent No. 5545553
;  GENERAL INFORMATION:
;    APPLICANT:  Gotschlich, Emil C.
;    TITLE OF INVENTION:  GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;    TITLE OF INVENTION:  OLIGOSACCHARIDES, AND GENES ENCODING THEM
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/312,387B
;      FILING DATE:  July 7, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-095
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  280 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-312-387B-6

  Query Match             70.9%;  Score 1036.5;  DB 1;  Length 280;
  Best Local Similarity   71.6%;  Pred. No. 3.4e-111;
  Matches  209;  Conservative   24;  Mismatches   34;  Indels   25;  Gaps    3;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||| ||||||||||||||||||||:||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||||||||:|||||||||||||||||||||||:|||| ||||:||
Db     61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||  |:  |  | :|  |:||||||||||||||||||:||||||||||||||||:
Db    121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||    |:|||||  |::|||| ||||||||  ||| ||| :| ||   |
Db    181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy    241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:: |: :: |:
Db    234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 7
```

```
US-08-683-426-6
; Sequence 6, Application US/08683426
; Patent No. 5705367
;  GENERAL INFORMATION:
;    APPLICANT:  Gotschlich, Emil C.
;    TITLE OF INVENTION:  GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;    TITLE OF INVENTION:  OLIGOSACCHARIDES, AND GENES ENCODING THEM
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/683,426
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/312,387
;      FILING DATE:  September 26, 1994
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-095B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  280 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-683-426-6

  Query Match             70.9%;  Score 1036.5;  DB 1;  Length 280;
  Best Local Similarity   71.6%;  Pred. No. 3.4e-111;
  Matches  209;  Conservative   24;  Mismatches   34;  Indels   25;  Gaps    3;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||| ||||||||||||||||||||:||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||||||||:|||||||||||||||||||||||:|||| ||||:||
Db     61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||  |:  |  | :|  |:||||||||||||||||||:||||||||||||||||:
Db    121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||    |:|||||  |::|||| ||||||||  ||| ||| :| ||   |
Db    181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy    241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:: |: :: |:
Db    234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 8
US-08-683-458-6
; Sequence 6, Application US/08683458
; Patent No. 5798233
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Gotschlich, Emil C.
;    TITLE OF INVENTION:  GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;    TITLE OF INVENTION:  OLIGOSACCHARIDES, AND GENES ENCODING THEM
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/683,458
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/312,387
;      FILING DATE:  September 26, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-095A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  280 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-683-458-6

  Query Match             70.9%;  Score 1036.5;  DB 1;  Length 280;
  Best Local Similarity   71.6%;  Pred. No. 3.4e-111;
  Matches  209;  Conservative   24;  Mismatches   34;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||| |||||||||||||||||||||:|||||||||||||||||
Db       1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||:|||||||||||||||||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:||||||||||||||||:||||||||||||||:
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||    |:||||| |::|||| |||||||  ||| ||| :| ||   |
Db     181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                 :||        | ||| | |||::|| :|:     |:: |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 9
US-08-878-360-6
; Sequence 6, Application US/08878360
; Patent No. 5945322
;  GENERAL INFORMATION:
;    APPLICANT:  Gotschlich, Emil C.
;    TITLE OF INVENTION:  GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;    TITLE OF INVENTION:  OLIGOSACCHARIDES, AND GENES ENCODING THEM
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,360
;      FILING DATE:  18-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/683,426
;      FILING DATE:
;      APPLICATION NUMBER:  08/312,387
;      FILING DATE:  September 26, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-095B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  280 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-878-360-6

  Query Match             70.9%;  Score 1036.5;  DB 2;  Length 280;
  Best Local Similarity   71.6%;  Pred. No. 3.4e-111;
  Matches  209;  Conservative   24;  Mismatches   34;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||| |||||||||||||||||||||:|||||||||||||||||
Db       1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||:|||||||||||||||||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:||||||||||||||||:||||||||||||||:
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||    |:||||| |::|||| |||||||  ||| ||| :| ||   |
Db     181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                 :||        | ||| | |||::|| :|:     |:: |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 10
US-08-478-140B-6
; Sequence 6, Application US/08478140B
; Patent No. 6127153
;  GENERAL INFORMATION:
;    APPLICANT:  JOHNSON, KARL F.
;    APPLICANT:  ROTH, STEPHEN
;    APPLICANT:  BUCZALA, STEPHANIE L.
```

```
;    TITLE OF INVENTION:  METHOD OF TRANSFERRING AT LEAST TWO
;    TITLE OF INVENTION:  SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
;    TITLE OF INVENTION:  POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
;    TITLE OF INVENTION:  POLYGLYCOSYLTRANSFERASE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,140B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Laura A. Coruzzi
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7188-017
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  280 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-478-140B-6

  Query Match             70.9%;  Score 1036.5;  DB 3;  Length 280;
  Best Local Similarity   71.6%;  Pred. No. 3.4e-111;
  Matches  209;  Conservative   24;  Mismatches   34;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||| ||||||||||||||||||||:||||||||||||||||||
Db-      1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||:|||||||||||||||||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:||||||||||||||||||:||||||||||||||||:
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||    |:||||| |::|||| ||||||||  ||| ||| :| ||   |
Db     181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:: |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 11
US-09-333-412-6
; Sequence 6, Application US/09333412
; Patent No. 6342382
;   GENERAL INFORMATION:
;        APPLICANT: Gotschlich, Emil C.
;        TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;                            OLIGOSACCHARIDES, AND GENES ENCODING THEM
;        NUMBER OF SEQUENCES: 12
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Klauber & Jackson
;             STREET: 411 Hackensack Avenue
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/333,412
;             FILING DATE: 15-Jun-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/312,387
;             FILING DATE: July 7, 1994
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jackson Esq., David A.
;             REGISTRATION NUMBER: 26,742
;             REFERENCE/DOCKET NUMBER: 600-1-095
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201 487-5800
;             TELEFAX: 201 343-1684
;             TELEX: 133521
;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 280 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-333-412-6

  Query Match             70.9%;  Score 1036.5;  DB 4;  Length 280;
  Best Local Similarity   71.6%;  Pred. No. 3.4e-111;
  Matches  209;  Conservative   24;  Mismatches   34;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||| ||||||||||||||||||||:||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||:|||||||||||||||||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:||||||||||||||||||:||||||||||||||||:
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||    |:||||| |::|||| ||||||||  ||| ||| :| ||   |
Db     181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:: |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 12
US-09-338-943-6
; Sequence 6, Application US/09338943
; Patent No. 6379933
;  GENERAL INFORMATION:
;    APPLICANT:  JOHNSON, KARL F.
;    APPLICANT:  ROTH, STEPHEN
;    APPLICANT:  BUCZALA, STEPHANIE L.
;    TITLE OF INVENTION:  METHOD OF TRANSFERRING AT LEAST TWO
;    TITLE OF INVENTION:  SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
;    TITLE OF INVENTION:  POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
;    TITLE OF INVENTION:  POLYGLYCOSYLTRANSFERASE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/338,943
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/478,140
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Laura A. Coruzzi
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7188-017
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  280 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-338-943-6

  Query Match             70.9%;  Score 1036.5;  DB 4;  Length 280;
  Best Local Similarity   71.6%;  Pred. No. 3.4e-111;
  Matches  209;  Conservative   24;  Mismatches   34;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||| |||||||||||||||||||:||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||:|||||||||||||||||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           ||||| |:  |  | :|  |:||||||||||||||||:|||||||||||||||:
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||    |:||||| |::|||| ||||||||  ||| ||| :| ||   |
Db     181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
              :||         | ||| | |||:|| :|:      |:: |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 13
US-08-701-191A-27
; Sequence 27, Application US/08701191A
; Patent No. 5942428
;  GENERAL INFORMATION:
;    APPLICANT:  Moosa Mohammadi, Joseph Schlessinger,
;    APPLICANT:  and Stevan R. Hubbard
;    TITLE OF INVENTION:  CRYSTALS OF THE TYROSINE KINASE DOMAIN
;    TITLE OF INVENTION:  OF NON-INSULIN RECEPTOR TYROSINE KINASE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/701,191A
;      FILING DATE:  August 21, 1996
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  227/088
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  304 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-701-191A-27

  Query Match              5.6%;  Score 82;  DB 2;  Length 304;
  Best Local Similarity   20.8%;  Pred. No. 0.39;
  Matches   35;  Conservative   39;  Mismatches   46;  Indels   48;  Gaps    8;

Qy      62 EKACFMSHAVLWEQALDEGVPYIA----VFED----DVLLGEGAEQFLAEDTWLQERFDP 113
           |:|  |   : :   : :|: |:|    |  |    ::|: || :  :::    :: ::
Db     133 ERALTMGDLISFAWQISQGMQYLAEMKLVHRDLAARNILVAEGRKMKISDFGLSRDVYEE 192

Qy     114 DSAFVVR----------LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKA 162
           || :| |           :|::| |: |: | |  :|   : ::  '     | |
Db     193 DS-YVKRSQGRIPVKWMAIESLFDHIYTTQSDVWSFGVLLWEIV------TLGG------ 239

Qy     163 MRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELH 210
               : :  :|||||   :|:  |:  :|         |  |::|::
Db     240 -----NPYPGIPPERL--FNLLKTGHRMER---------PDNCSEEMY 271


RESULT 14
US-09-134-001C-3799
; Sequence 3799, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3799
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3799
```

Query Match 5.6%; Score 81.5; DB 4; Length 337;
Best Local Similarity 22.1%; Pred. No. 0.52;
Matches 50; Conservative 39; Mismatches 64; Indels 73; Gaps 13;

```
Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFF--DALMPSERLERAMAELVPG---LSAH 55
          : :|:| :     | :   | |    |||  :  ||:||   |    |::| |   :  :
Db    150 IMHHLIEVDPVLSRVSVHLANF--LEIPFDVYPQDAIMP---LPLEEADVVIGDFPIGYY 204

Qy     56 P-------YLSGVEKACFMSHAVLWEQALD----EGVPYIAV----FEDDVLLGEGAEQF 100
          |       | ||    || :| ||:::     |  :: |    |||| :  :  | |
Db    205 PLDERSREMKLGFEKGHSYSHHLLIEQSINALKGAGYAFLVVPSHLFEDDKV--KQLENF 262

Qy    101 LAEDTWLQERFDPDSAFVVRLETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISR 160
          :| :| :|       ||:   :|:| :         :   ::  :|: :  |
Db    263 IATETEMQ-------AFLNLPKTLFKN---------EKARKSILILQKKKSG-------- 298

Qy    161 KAMRFFLDRFAVLPPERLHPVDLMM-----FGNPDDREGMPVCQLN 201
                          ||::::     | ||   :|  : :||
Db    299 ----------------ETRPVEVLLANIPDFKNPQQFQGF-ISELN 327
```

RESULT 15
US-09-564-805-236
; Sequence 236, Application US/09564805
; Patent No. 6333403
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.F.
; APPLICANT: Simard, Jacques
; APPLICANT: Rommens, Johanna M.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
; TITLE OF INVENTION: Gene and a Paralog and Orthologous Genes
; FILE REFERENCE: 2318-258
; CURRENT APPLICATION NUMBER: US/09/564,805
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/107,468
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: 09/434,382
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 236
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Synechocystis sp.
US-09-564-805-236

Query Match 5.5%; Score 81; DB 4; Length 554;
Best Local Similarity 21.5%; Pred. No. 1.3;
Matches 64; Conservative 37; Mismatches 99; Indels 98; Gaps 17;

```
Qy     22 FGSRGIPF------QFFDA-----------LMPSERLERAMAELVPGLSAHPYLSG---- 60
          :|:| :|       ||  |           |:|   |  |  |::  |  |   :|
Db    191 YGNRRLPHRRQQEKQFIQAIETVLAKGRNILLPVPPLGLAQ-EILKLLRTHHQFTGRQVN 249

Qy     61 ------VEKAC---------------FMSHAVL-WEQALDEGVPYIAVFEDDVLLGE-- 95
              | : |               |  |  | |:   |:  |::    ||   ||
Db    250 LWAGESVARGCDAYQGIIDHLPDNVRNFAQHQPLFWD---DKVYPHLRPLTDD--QGELS 304

Qy     96 -GAEQFLAEDTWLQERFDPDSAFVVRLETMFM-HVLTSPSGVADYGGR---AFPLLESEH 150
           |   :   ||    | |  | :  | |:||  :|| || : ::  :    ||  | |
Db    305 LSAPSIVITTTW--PAFWPSPAALPGLWTVFMPQLLTLPSCLVNFAWQDLEEFPKYELED 362

Qy    151 CGTAGYIISRKAMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVC-------QLN-- 201
            | :   |   :   :    | |: |    : : | | | | :           ||:
Db    363 YLLADHSDGRNTTQLIHN----LRPQHL----VFVHGQPSDIEDLTSLEELQSRYQLHSP 414

Qy    202 --------------------PALCAQELHYAKFHDQNSALGSLIEH-DRRLNRKQQW 237
                             | :   |:|  : : |   || :: | | ::    :|
Db    415 AAGNAVALPIGDRFVQPTPPPPQIYEGEIHELEPNKQIHHLGEVVIHLDGQILENSRW 472
```

Search completed: May 27, 2003, 18:18:46
Job time : 30 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 27, 2003, 18:17:25 ; Search time 23 Seconds
(without alignments)
1203.008 Million cell updates/sec

Title: US-10-007-267A-8
Perfect score: 1461
Sequence: 1 MQNHVISLASAAERRAHIAA.....REKRRQRREQLIGKIIVPFQ 279

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1461	100.0	279	12	US-10-007-267-8	Sequence 8, Appli
2	1288	88.2	275	10	US-09-211-691-2	Sequence 2, Appli
3	1036.5	70.9	280	12	US-10-007-267-6	Sequence 6, Appli
4	98.5	6.7	273	9	US-10-218-381-1	Sequence 1, Appli
5	95.5	6.5	389	9	US-10-235-056-7	Sequence 7, Appli
6	82.5	5.6	273	9	US-09-895-913A-186	Sequence 186, App
7	82.5	5.6	366	9	US-10-235-056-9	Sequence 9, Appli
8	81	5.5	554	9	US-09-988-626-236	Sequence 236, App
9	81	5.5	554	9	US-09-988-687-236	Sequence 236, App
10	80	5.5	222	10	US-09-864-761-48086	Sequence 48086, A
11	80	5.5	285	9	US-10-253-007-54	Sequence 54, Appl
12	80	5.5	350	9	US-10-235-056-11	Sequence 11, Appl
13	80	5.5	363	9	US-10-235-056-1	Sequence 1, Appli
14	80	5.5	363	10	US-09-849-031A-1	Sequence 1, Appli
15	80	5.5	363	10	US-09-849-562A-1	Sequence 1, Appli
16	78	5.3	302	10	US-09-815-242-5371	Sequence 5371, Ap
17	78	5.3	325	10	US-09-815-242-12608	Sequence 12608, A
18	77	5.3	265	9	US-09-454-279-22	Sequence 22, Appl
19	77	5.3	570	10	US-09-764-853-551	Sequence 551, App
20	76	5.2	362	9	US-10-235-056-3	Sequence 3, Appli
21	76	5.2	755	9	US-09-738-626-4773	Sequence 4773, Ap
22	76	5.2	893	10	US-09-916-790-5	Sequence 5, Appli
23	75	5.1	628	9	US-10-097-340-208	Sequence 208, App
24	74.5	5.1	275	10	US-09-841-132-443	Sequence 443, App
25	74.5	5.1	1443	10	US-09-815-242-5004	Sequence 5004, Ap
26	74.5	5.1	1451	10	US-09-815-242-10734	Sequence 10734, A
27	74	5.1	352	10	US-09-739-451-4	Sequence 4, Appli
28	74	5.1	353	9	US-10-097-065-137	Sequence 137, App
29	74	5.1	353	10	US-09-739-451-12	Sequence 12, Appl
30	73.5	5.0	3034	10	US-09-737-149-25	Sequence 25, Appl
31	73.5	5.0	3034	10	US-09-737-149-30	Sequence 30, Appl
32	73	5.0	738	9	US-10-235-521-1	Sequence 1, Appli
33	72.5	5.0	388	9	US-10-235-056-17	Sequence 17, Appl
34	72	4.9	3472	9	US-10-027-806-4	Sequence 4, Appli
35	72	4.9	3472	9	US-10-034-623-4	Sequence 4, Appli
36	72	4.9	3472	9	US-10-027-801-4	Sequence 4, Appli
37	71.5	4.9	254	10	US-09-796-149-4	Sequence 4, Appli
38	71.5	4.9	261	9	US-10-166-087-34	Sequence 34, Appl
39	71.5	4.9	348	9	US-09-738-626-5823	Sequence 5823, Ap
40	71.5	4.9	436	9	US-09-764-868-826	Sequence 826, App
41	71.5	4.9	792	9	US-10-029-115-4	Sequence 4, Appli
42	71.5	4.9	966	10	US-09-978-698-2	Sequence 2, Appli
43	71.5	4.9	1244	9	US-09-789-390-13	Sequence 13, Appl
44	71.5	4.9	1273	9	US-09-789-390-11	Sequence 11, Appl
45	71.5	4.9	1295	9	US-09-789-390-30	Sequence 30, Appl

ALIGNMENTS

```
RESULT 1
US-10-007-267-8
; Sequence 8, Application US/10007267
; Patent No. US20020127682A1
;   GENERAL INFORMATION:
;        APPLICANT: Gotschlich, Emil C.
;        TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;                            OLIGOSACCHARIDES, AND GENES ENCODING THEM
;        NUMBER OF SEQUENCES: 12
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Klauber & Jackson
;             STREET: 411 Hackensack Avenue
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/007,267
;             FILING DATE: 03-Dec-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/333,412
;             FILING DATE: 15-Jun-1999
;             APPLICATION NUMBER: 08/312,387
;             FILING DATE: July 7, 1994
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jackson Esq., David A.
;             REGISTRATION NUMBER: 26,742
;             REFERENCE/DOCKET NUMBER: 600-1-095
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201 487-5800
;             TELEFAX: 201 343-1684
;             TELEX: 133521
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 279 amino acids
;             TYPE: amino acid
```

```
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-007-267-8

  Query Match             100.0%;  Score 1461;  DB 12;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-143;
  Matches  279;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          |||||||||||||||||||||||||||||||||||||||
Db    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279


RESULT 2
US-09-211-691-2
; Sequence 2, Application US/09211691
; Patent No. US20020034805A1
; GENERAL INFORMATION:
;  APPLICANT:  Gilbert, Michel
;  APPLICANT:  Young, N. Martin
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION:  Fusion Proteins for Use in Enzymatic Synthesis of
;  TITLE OF INVENTION:  Oligosaccharides
;  FILE REFERENCE:  019957-012910US
;  CURRENT APPLICATION NUMBER: US/09/211,691
;  CURRENT FILING DATE:  1998-12-14
;  PRIOR APPLICATION NUMBER:  US 60/069,443
;  PRIOR FILING DATE:  1997-12-15
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 275
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-211-691-2

  Query Match             88.2%;  Score 1288;  DB 10;  Length 275;
  Best Local Similarity   90.0%;  Pred. No. 4.6e-125;
  Matches  251;  Conservative    7;  Mismatches   17;  Indels    4;  Gaps    1;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||| |||  ||||||||||||||||:|||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGRHGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:||||||:||| |||||||||| |:||||| |||||||||:||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYITVFEDDVLLGEGEEKFLAEDAWLQERFDPDTAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||||||||||||||| |||||||||| |||||||||||||||||| |||| |||
Db    121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMRFFLDRFAALPPEGLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          |||||| : |||||||||||||||||||||||||||||||||||||| |||||| |||
Db    181 VDLMMFSDFFDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRLLNRKQQRRDS 240

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          |||||||||||||||| |||||||||||     |||||
Db    241 PANTFKHRLIRALTKISREREKRRQRREQ----FIVPFQ 275


RESULT 3
US-10-007-267-6
; Sequence 6, Application US/10007267
; Patent No. US20020127682A1
;  GENERAL INFORMATION:
;       APPLICANT: Gotschlich, Emil C.
;       TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;                           OLIGOSACCHARIDES, AND GENES ENCODING THEM
;       NUMBER OF SEQUENCES: 12
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Klauber & Jackson
;            STREET: 411 Hackensack Avenue
;            CITY: Hackensack
;            STATE: New Jersey
;            COUNTRY: USA
;            ZIP: 07601
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/007,267
;            FILING DATE: 03-Dec-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/333,412
;            FILING DATE: 15-Jun-1999
;            APPLICATION NUMBER: 08/312,387
;            FILING DATE: July 7, 1994
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Jackson Esq., David A.
;            REGISTRATION NUMBER: 26,742
;            REFERENCE/DOCKET NUMBER: 600-1-095
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 201 487-5800
;            TELEFAX: 201 343-1684
;            TELEX: 133521
;  INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 280 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-007-267-6

  Query Match             70.9%;  Score 1036.5;  DB 12;  Length 280;
  Best Local Similarity   71.6%;  Pred. No. 4.7e-99;
  Matches  209;  Conservative   24;  Mismatches   34;  Indels   25;  Gaps    3;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||| |||||||||||||||||||:|||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||||||||:|||||||||||||||||||||||||:|||| ||||:||
Db     61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||  |:  |  | :|  |:|||||||||||||||||:|||||||||||||||:
Db    121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||    |:|||| |::|||| ||||||||  ||| ||| :| ||   |
Db    181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233
```

```
Qy    241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
          :||        | ||| | |||::|| :|:     |:: |: :: |:
Db    234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 4
US-10-218-381-1
; Sequence 1, Application US/10218381
; Publication No. US20030013175A1
; GENERAL INFORMATION:
;  APPLICANT: Satoshi KOIZUMI.
;  APPLICANT:  Tetsuo ENDO
;  APPLICANT:  Kazuhiko TABATA
;  APPLICANT:  Akio OZAKI
;  TITLE OF INVENTION: GLYCOSYLTRANSFERASE AND DNA ENCODING THE SAME
;  FILE REFERENCE: 2139.18
;  CURRENT APPLICATION NUMBER: US/10/218,381
;  CURRENT FILING DATE:  2002-08-15
;  PRIOR APPLICATION NUMBER: US/09/601,519
;  PRIOR FILING DATE: 2000-08-02
;  PRIOR APPLICATION NUMBER: PCT/JP99/00321
;  PRIOR FILING DATE: 1999-01-27
;  PRIOR APPLICATION NUMBER: JP 10-023389
;  PRIOR FILING DATE: 1998-02-04
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 273
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-10-218-381-1

  Query Match             6.7%;  Score 98.5;  DB 9;  Length 273;
  Best Local Similarity   23.9%;  Pred. No. 0.042;
  Matches   50;  Conservative   24;  Mismatches   54;  Indels   81;  Gaps   12;

Qy     30 QFFDALMPSERLERAMAELVPGLSAHPY---------------LSGVEK---------- 63
          | |||:  |:  |  :  ||     |||                :| | |
Db     38 QIFDAIY-SKTFEGGLHPLVK-KHLHPYFITQNIKDMGITTNLISEVSKFYYALKYHAKF 95

Qy     64 ------ACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAF 117
                |: ||  |||: ::     | : |||: | |  :: |    :|::
Db     96 MSLGELGCYASHYSLWEKCIELN-EAICILEDDITLKEDFKEGL---DFLEKHI--QELG 149

Qy    118 VVRLETMFMHVLTSPSGVADYGGRAFPLLESEH---------------CGTAGYIISRKA 162
           |||    ||:|  |:  ::      ||    |                || ||:|: |
Db    150 YVRL----MHLLYDPNVKSE------PLNHKNHEIQERVGIIKAYSHGVGTQGYVITPKI 199

Qy    163 MRFFLDRFAVLPPERLH------PVDLMM 185
          : |          : |      ||| :|
Db    200 AKVF----------KKHSRKWVVPVDTIM 218


RESULT 5
US-10-235-056-7
; Sequence 7, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
;  APPLICANT: Canfield, William M.
;  APPLICANT:  Cummings, Richard D.
;  APPLICANT:  Ju, Tongzhong
;  TITLE OF INVENTION: CORE 1  3-GALACTOSYL TRANSFERASES
;  FILE REFERENCE: 5820.630
;  CURRENT APPLICATION NUMBER: US/10/235,056
;  CURRENT FILING DATE:  2002-09-04
;  PRIOR APPLICATION NUMBER: US 09/461,321
;  PRIOR FILING DATE: 1999-12-15
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 389
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
US-10-235-056-7

  Query Match             6.5%;  Score 95.5;  DB 9;  Length 389;
  Best Local Similarity   23.1%;  Pred. No. 0.14;
  Matches   42;  Conservative   26;  Mismatches   63;  Indels   51;  Gaps    8;

Qy      2 QNHVISLASAAERRAHIAATFGSRGIPFQFF----DALMPSERLERAMAELVPGLSAHPY 57
          |||        :|  |: ||:  |   : |     || :|:  |  :          |
Db    115 QNH-------DKRAKHVKATWAKRCNKYVFMSSEEDAELPAINLNVS--------EGRDY 159

Qy     58 LSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAF 117
          |    |  |     :::  |::                    : ||  |         |  :
Db    160 LWAKTKGAF---KYIYDHHLND----------------YDWFLKAD---------DDTY 190

Qy    118 VVRLETMFMHVLTSPSGVADYGGRAFPLLE-SEHCGTAGYIISRKAMRFFLDRFAVLPPE 176
          ||    || :  ||      :| :  |  :   | | |||::||:|:: |::    || :
Db    191 VVMENLRFMLLAHSPDEPIHFGCKFKPFTQGGYHSGGAGYVLSREALKKFIE--VALPDK 248

Qy    177 RL 178
          |
Db    249 SL 250


RESULT 6
US-09-895-913A-186
; Sequence 186, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
;  APPLICANT: Kleanthous, Harold
;  APPLICANT:  Al-Garawi, Amal
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Tomb, Jean Francois
;  APPLICANT:  Oomen, Raymond P.
;  TITLE OF INVENTION: Identification of Polynucleotides
;  TITLE OF INVENTION:  Encoding No. US20020160456A1el Helicobacter Polypeptides in t
;  TITLE OF INVENTION:  Genome
;  FILE REFERENCE: 06132/043002
;  CURRENT APPLICATION NUMBER: US/09/895,913A
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 08/881,227
;  PRIOR FILING DATE: 1997-06-24
;  NUMBER OF SEQ ID NOS: 368
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 186
;   LENGTH: 273
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-09-895-913A-186

  Query Match             5.6%;  Score 82.5;  DB 9;  Length 273;
  Best Local Similarity   21.4%;  Pred. No. 1.9;
  Matches   45;  Conservative   33;  Mismatches   73;  Indels   59;  Gaps    8;

Qy     41 LERAMAELVPGLSAHPYLSGVEKACFMSHAVLWEQALD---EGVPYIAVFEDDVLLGEGA 97
          ||||: |                    : |  |:|: |:   :   ::   :: |  ||
Db      7 LERALKE-----------------DLGHGDLFERVLEKDFKATAFVRAKQEGVFSGEKY 48

Qy     98 EQFLAEDTWLQ--------ERFDP---------DSAFVVRLETMFMHVLTSPSGVADYGG 140
             | | | ::        ||| |         | : ::::|   :::|   ||:|
Db     49 ALELLEMTGIECVQTIKDKERFKPKDALMEIRGDFSMLLKVERTLLNLLQHSSGIATLTS 108

Qy    141 RAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVCQL 200
          |    | |          :|  :| | ::::||              |  : | |     |
Db    109 RFVEALNSHKVRLLDTRKTRPLLRIF-EKYSVLNG----------GASNHRLG-----L 151

Qy    201 NPALCAQELHYAKFHDQNSALGSLIEHDRR 230
          : ||  :: |     |    | | : | |:
Db    152 DDALMLKDTHLRHVKD----LKSFLTHARK 177


RESULT 7
US-10-235-056-9
; Sequence 9, Application US/10235056
```

```
; Publication No. US20030059915A1
; GENERAL INFORMATION:
;  APPLICANT: Canfield, William M.
;  APPLICANT:  Cummings, Richard D.
;  APPLICANT:  Ju, Tongzhong
;  TITLE OF INVENTION: CORE 1  3-GALACTOSYL TRANSFERASES
;  FILE REFERENCE: 5820.630
;  CURRENT APPLICATION NUMBER: US/10/235,056
;  CURRENT FILING DATE:  2002-09-04
;  PRIOR APPLICATION NUMBER: US 09/461,321
;  PRIOR FILING DATE: 1999-12-15
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
;   LENGTH: 366
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-235-056-9

  Query Match             5.6%;  Score 82.5;  DB 9;  Length 366;
  Best Local Similarity   22.9%;  Pred. No. 2.8;
  Matches   44;  Conservative   23;  Mismatches   48;  Indels   77;  Gaps   11;

Qy      63 KAC----FMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFD------ 112
           | |    |||         || :| : :.  ||  ||| |   |:   ::| |
Db     119 KRCNILLFMS------SGADEELPTVKL---DV--GEGRENLWAK---VKEAFKYVYHHH 164

Qy     113 ----------PDSAFVVRLETMFMHVLTSPSGVADYGGRAFPLLESEH-CGTAGYIISRK 161
                    |  : |     :|    :|    :| :  | ::  :  | ||||:||:
Db     165 YNDADFFYKADDDTYAVIENMRYMLYPYNPETPVHFGFKFKPFVKQGYMSGGAGYILSRE 224

Qy     162 AMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSAL 221
           |:|    || |                     ||:|    || :|
Db     225 ALR----RFVV-------------------EGIP----NPKMCLP-------------- 242

Qy     222 GSLIEHDRRLNR 233
           |:::  |  : |
Db     243 GTVVNEDIEIGR 254


RESULT 8
US-09-988-626-236
; Sequence 236, Application US/09988626
; Publication No. US20030044959A1
; GENERAL INFORMATION:
;  APPLICANT: Tavtigian, Sean V.
;  APPLICANT:  Teng, David H.F.
;  APPLICANT:  Simard, Jacques
;  APPLICANT:  Rommens, Johanna M.
;  APPLICANT:  Myriad Genetics, Inc.
;  TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
;  TITLE OF INVENTION:  Gene and a Paralog and Orthologous Genes
;  FILE REFERENCE: 2318-258
;  CURRENT APPLICATION NUMBER: US/09/988,626
;  CURRENT FILING DATE:  2001-11-20
;  PRIOR APPLICATION NUMBER: 09/564,805
;  PRIOR FILING DATE: 2000-05-05
;  PRIOR APPLICATION NUMBER: US 60/107,468
;  PRIOR FILING DATE: 1998-11-06
;  PRIOR APPLICATION NUMBER: 09/434,382
;  PRIOR FILING DATE: 1999-11-05
;  NUMBER OF SEQ ID NOS: 240
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 236
;   LENGTH: 554
;   TYPE: PRT
;   ORGANISM: Synechocystis sp.
US-09-988-626-236

  Query Match             5.5%;  Score 81;  DB 9;  Length 554;
  Best Local Similarity   21.5%;  Pred. No. 7.1;
  Matches   64;  Conservative   37;  Mismatches   99;  Indels   98;  Gaps   17;

Qy      22 FGSRGIPF------QFFDA-----------LMPSERLERAMAELVPGLSAHPYLSG---- 60
           :|:| :|       ||  |            |:|   |  |  |::  |  |   :|
Db     191 YGNRRLPHRRQQEKQFIQAIETVLAKGRNILLPVPPLGLAQ-EILKLLRTHHQFTGRQVN 249

Qy      61 ------VEKAC----------------FMSHAVL-WEQALDEGVPYIAVFEDDVLLGE-- 95
                | : |                |  |  | |:   |:  |::    ||   ||
Db     250 LWAGESVARGCDAYQGIIDHLPDNVRNFAQHQPLFWD---DKVYPHLRPLTDD--QGELS 304

Qy      96 -GAEQFLAEDTWLQERFDPDSAFVVRLETMFM-HVLTSPSGVADYGGR---AFPLLESEH 150
            |   :  ||    | |  | :  | |:||  :|| || : ::  :    ||  | |
Db     305 LSAPSIVITTTW--PAFWPSPAALPGLWTVFMPQLLTLPSCLVNFAWQDLEEFPKYELED 362

Qy     151 CGTAGYIISRKAMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVC-------QLN-- 201
           | :   |   :   :    | |: |    : : | | | :           ||:
Db     363 YLLADHSDGRNTTQLIHN----LRPQHL----VFVHGQPSDIEDLTSLEELQSRYQLHSP 414

Qy     202 ---------------------PALCAQELHYAKFHDQNSALGSLIEH-DRRLNRKQQW 237
                               | :   |:|  : : |   || :: | | ::    :|
Db     415 AAGNAVALPIGDRFVQPTPPPPQIYEGEIHELEPNKQIHHLGEVVIHLDGQILENSRW 472


RESULT 9
US-09-988-687-236
; Sequence 236, Application US/09988687
; Publication No. US20030045704A1
; GENERAL INFORMATION:
;  APPLICANT: Tavtigian, Sean V.
;  APPLICANT:  Teng, David H.F.
;  APPLICANT:  Simard, Jacques
;  APPLICANT:  Rommens, Johanna M.
;  APPLICANT:  Myriad Genetics, Inc.
;  TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
;  TITLE OF INVENTION:  Gene and a Paralog and Orthologous Genes
;  FILE REFERENCE: 2318-258
;  CURRENT APPLICATION NUMBER: US/09/988,687
;  CURRENT FILING DATE:  2001-11-20
;  PRIOR APPLICATION NUMBER: 09/564,805
;  PRIOR FILING DATE: 2000-05-05
;  PRIOR APPLICATION NUMBER: US 60/107,468
;  PRIOR FILING DATE: 1998-11-06
;  PRIOR APPLICATION NUMBER: 09/434,382
;  PRIOR FILING DATE: 1999-11-05
;  NUMBER OF SEQ ID NOS: 240
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 236
;   LENGTH: 554
;   TYPE: PRT
;   ORGANISM: Synechocystis sp.
US-09-988-687-236

  Query Match             5.5%;  Score 81;  DB 9;  Length 554;
  Best Local Similarity   21.5%;  Pred. No. 7.1;
  Matches   64;  Conservative   37;  Mismatches   99;  Indels   98;  Gaps   17;

Qy      22 FGSRGIPF------QFFDA-----------LMPSERLERAMAELVPGLSAHPYLSG---- 60
           :|:| :|       ||  |            |:|   |  |  |::  |  |   :|
Db     191 YGNRRLPHRRQQEKQFIQAIETVLAKGRNILLPVPPLGLAQ-EILKLLRTHHQFTGRQVN 249

Qy      61 ------VEKAC----------------FMSHAVL-WEQALDEGVPYIAVFEDDVLLGE-- 95
                | : |                |  |  | |:   |:  |::    ||   ||
Db     250 LWAGESVARGCDAYQGIIDHLPDNVRNFAQHQPLFWD---DKVYPHLRPLTDD--QGELS 304

Qy      96 -GAEQFLAEDTWLQERFDPDSAFVVRLETMFM-HVLTSPSGVADYGGR---AFPLLESEH 150
            |   :  ||    | |  | :  | |:||  :|| || : ::  :    ||  | |
Db     305 LSAPSIVITTTW--PAFWPSPAALPGLWTVFMPQLLTLPSCLVNFAWQDLEEFPKYELED 362

Qy     151 CGTAGYIISRKAMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVC-------QLN-- 201
           | :   |   :   :    | |: |    : : | | | :           ||:
Db     363 YLLADHSDGRNTTQLIHN----LRPQHL----VFVHGQPSDIEDLTSLEELQSRYQLHSP 414

Qy     202 ---------------------PALCAQELHYAKFHDQNSALGSLIEH-DRRLNRKQQW 237
                               | :   |:|  : : |   || :: | | ::    :|
Db     415 AAGNAVALPIGDRFVQPTPPPPQIYEGEIHELEPNKQIHHLGEVVIHLDGQILENSRW 472
```

```
RESULT 10
US-09-864-761-48086
; Sequence 48086, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48086
;   LENGTH: 222
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005532.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;   OTHER INFORMATION: SWISSPROT HIT: Q50950, EVALUE 4.90e-01
;   OTHER INFORMATION: EST_HUMAN HIT: BE336820.1, EVALUE 1.00e-112
US-09-864-761-48086

  Query Match             5.5%;  Score 80;  DB 10;  Length 222;
  Best Local Similarity   26.1%;  Pred. No. 2.6;
  Matches   23;  Conservative   14;  Mismatches   19;  Indels   32;  Gaps    4;

Qy      95 EGAEQFLAED----------TWLQERFDPDSAFVVRLETMFMHVLTSPSGVADYGGRAF- 143
           | |: ||  |          ||  ::||:                        | || |
Db      86 EDADWFLKADDDTYVILDNLRWLLSKYDPEEPI--------------------YFGRRFK 125

Qy     144 PLLESEH-CGTAGYIISRKAMRFFLDRF 170
           | ::  :  | |||::|::|:: |:| |
Db     126 PYVKQGYMSGGAGYVLSKEALKRFVDAF 153


RESULT 11
US-10-253-007-54
; Sequence 54, Application US/10253007
; Publication No. US20030088073A1
; GENERAL INFORMATION:
;  APPLICANT: Benfey et al.
;  TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 5914-074-999
;  CURRENT APPLICATION NUMBER: US/10/253,007
;  CURRENT FILING DATE:  2002-09-23
;  PRIOR APPLICATION NUMBER: US/09/186,188
;  PRIOR FILING DATE: 1998-11-05
;  PRIOR APPLICATION NUMBER: 08/842,445
;  PRIOR FILING DATE: 1997-04-24
;  PRIOR APPLICATION NUMBER: 08/638,617
;  PRIOR FILING DATE: 1996-04-26
;  NUMBER OF SEQ ID NOS: 79
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 54
;   LENGTH: 285
;   TYPE: PRT
;   ORGANISM: Plant
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(285)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-10-253-007-54

  Query Match             5.5%;  Score 80;  DB 9;  Length 285;
  Best Local Similarity   22.8%;  Pred. No. 3.6;
  Matches   61;  Conservative   35;  Mismatches   88;  Indels   84;  Gaps   16;

Qy       8 LASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSA---HPYLSGVEKA 64
           |:   || | :|    | |:||:| ||:|   :::|    | || :     ||:
Db      63 LSLVGERLATLAQ---SCGVPFEFHDAIMSGCKVQREHLGLEPGFAVVVNFPYV------ 113

Qy      65 CFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVRLETM 124
             : |         || |  :  : | :|               | :   |    :|  |:
Db     114 --LHHMP------DESVS-VEKYRDRLL-------------HLIKSLSPKLVTLVEQES- 150

Qy     125 FMHVLTSP-----SGVADYGGRAFPLLES-------------EHCGTAGYIISRKAMRFF 166
             :  |||          ||    |  :::             :||      ::|  :
Db     151 --NTNTSPLVSRFVETLDYYTAMFESIDAARPRDDKQRISAEQHC------VARDIVNMI 202

Qy     167 -------LDRFAVLPPERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNS 219
                  ::|  ||    |   | :|| |      | ||   : |  | |:   | :|:|
Db     203 ACEESERVERHEVLGKWR---VRMMMAG----FTGWPV-STSAAFAASEM--LKAYDKNY 252

Qy     220 ALGS------LIEHDRRLNRKQQWRDSP 241
             ||      |    | :     |: :|
Db     253 KLGGHEGALYLFWKRRPMATCSVWKPNP 280


RESULT 12
US-10-235-056-11
; Sequence 11, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
;  APPLICANT: Canfield, William M.
;  APPLICANT:  Cummings, Richard D.
;  APPLICANT:  Ju, Tongzhong
;  TITLE OF INVENTION: CORE 1  3-GALACTOSYL TRANSFERASES
;  FILE REFERENCE: 5820.630
;  CURRENT APPLICATION NUMBER: US/10/235,056
;  CURRENT FILING DATE:  2002-09-04
;  PRIOR APPLICATION NUMBER: US 09/461,321
```

```
;  PRIOR FILING DATE: 1999-12-15
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
;   LENGTH: 350
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-235-056-11

  Query Match             5.5%;  Score 80;  DB 9;  Length 350;
  Best Local Similarity   26.1%;  Pred. No. 4.8;
  Matches   23;  Conservative   14;  Mismatches   19;  Indels   32;  Gaps    4;

Qy      95 EGAEQFLAED----------TWLQERFDPDSAFVVRLETMFMHVLTSPSGVADYGGRAF- 143
           | |: ||  |           ||  ::||:                    | || |
Db     147 EDADWFLKADDDTYVILDNLRWLLSKYDPEEPI--------------------YFGRRFK 186

Qy     144 PLLESEH-CGTAGYIISRKAMRFFLDRF 170
           | ::  :  | |||::|::|:: |:| |
Db     187 PYVKQGYMSGGAGYVLSKEALKRFVDAF 214


RESULT 13
US-10-235-056-1
; Sequence 1, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
;  APPLICANT: Canfield, William M.
;  APPLICANT:  Cummings, Richard D.
;  APPLICANT:  Ju, Tongzhong
;  TITLE OF INVENTION: CORE 1  3-GALACTOSYL TRANSFERASES
;  FILE REFERENCE: 5820.630
;  CURRENT APPLICATION NUMBER: US/10/235,056
;  CURRENT FILING DATE:  2002-09-04
;  PRIOR APPLICATION NUMBER: US 09/461,321
;  PRIOR FILING DATE: 1999-12-15
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 363
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-235-056-1

  Query Match             5.5%;  Score 80;  DB 9;  Length 363;
  Best Local Similarity   26.1%;  Pred. No. 5.1;
  Matches   23;  Conservative   14;  Mismatches   19;  Indels   32;  Gaps    4;

Qy      95 EGAEQFLAED----------TWLQERFDPDSAFVVRLETMFMHVLTSPSGVADYGGRAF- 143
           | |: ||  |           ||  ::||:                    | || |
Db     160 EDADWFLKADDDTYVILDNLRWLLSKYDPEEPI--------------------YFGRRFK 199

Qy     144 PLLESEH-CGTAGYIISRKAMRFFLDRF 170
           | ::  :  | |||::|::|:: |:| |
Db     200 PYVKQGYMSGGAGYVLSKEALKRFVDAF 227


RESULT 14
US-09-849-031A-1
; Sequence 1, Application US/09849031A
; Patent No. US20020026033A1
; GENERAL INFORMATION:
;  APPLICANT: Cummings, Richard D.
;  APPLICANT:  McEver, Rodger P.
;  TITLE OF INVENTION: synthetic glycosulfopeptides and methods of synthesis thereof
;  FILE REFERENCE: 5820.593
;  CURRENT APPLICATION NUMBER: US/09/849,031A
;  CURRENT FILING DATE:  2001-05-04
;  PRIOR APPLICATION NUMBER: 09/334,013
;  PRIOR FILING DATE: 1999-06-15
;  NUMBER OF SEQ ID NOS: 40
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 363
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-849-031A-1

  Query Match             5.5%;  Score 80;  DB 10;  Length 363;
  Best Local Similarity   26.1%;  Pred. No. 5.1;
  Matches   23;  Conservative   14;  Mismatches   19;  Indels   32;  Gaps    4;

Qy      95 EGAEQFLAED----------TWLQERFDPDSAFVVRLETMFMHVLTSPSGVADYGGRAF- 143
           | |: ||  |           ||  ::||:                    | || |
Db     160 EDADWFLKADDDTYVILDNLRWLLSKYDPEEPI--------------------YFGRRFK 199

Qy     144 PLLESEH-CGTAGYIISRKAMRFFLDRF 170
           | ::  :  | |||::|::|:: |:| |
Db     200 PYVKQGYMSGGAGYVLSKEALKRFVDAF 227


RESULT 15
US-09-849-562A-1
; Sequence 1, Application US/09849562A
; Patent No. US20020042102A1
; GENERAL INFORMATION:
;  APPLICANT: Cummings, Richard D.
;  APPLICANT:  McEver, Rodger P.
;  TITLE OF INVENTION: synthetic glycosulfopeptides and methods of synthesis thereof
;  FILE REFERENCE: 5820.594
;  CURRENT APPLICATION NUMBER: US/09/849,562A
;  CURRENT FILING DATE:  2001-05-04
;  PRIOR APPLICATION NUMBER: 09/334,013
;  PRIOR FILING DATE: 1999-06-15
;  NUMBER OF SEQ ID NOS: 40
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 363
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-849-562A-1

  Query Match             5.5%;  Score 80;  DB 10;  Length 363;
  Best Local Similarity   26.1%;  Pred. No. 5.1;
  Matches   23;  Conservative   14;  Mismatches   19;  Indels   32;  Gaps    4;

Qy      95 EGAEQFLAED----------TWLQERFDPDSAFVVRLETMFMHVLTSPSGVADYGGRAF- 143
           | |: ||  |           ||  ::||:                    | || |
Db     160 EDADWFLKADDDTYVILDNLRWLLSKYDPEEPI--------------------YFGRRFK 199

Qy     144 PLLESEH-CGTAGYIISRKAMRFFLDRF 170
           | ::  :  | |||::|::|:: |:| |
Db     200 PYVKQGYMSGGAGYVLSKEALKRFVDAF 227


Search completed: May 27, 2003, 18:23:31
Job time : 25 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 27, 2003, 18:12:59 ; Search time 45 Seconds
(without alignments)
596.033 Million cell updates/sec

Title: US-10-007-267A-8
Perfect score: 1461
Sequence: 1 MQNHVISLASAAERRAHIAA.....REKRRQRREQLIGKIIVPFQ 279

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1329	91.0	279	2	A81971	lacto-N-neotetraos
2	1293	88.5	275	2	C81027	lacto-N-neotetraos
3	1288	88.2	275	2	S70814	glycosyl transfera
4	997.5	68.3	268	2	B81971	lacto-N-neotetraos
5	928.5	63.6	280	2	A81027	lacto-N-neotetraos
6	881.5	60.3	276	2	S70815	glycosyl transfera
7	461	31.6	266	2	S71025	lipopolysaccharide
8	418	28.6	302	1	A64077	lipopolysaccharide
9	271	18.5	282	2	F64091	lipopolysaccharide
10	231.5	15.8	268	2	AD3304	lipooligosaccharid
11	128	8.8	332	2	E71916	probable lipopolys
12	116.5	8.0	721	2	AH3417	lpsA protein [impo
13	107	7.3	738	2	T00343	hypothetical prote
14	100.5	6.9	267	2	E64175	hypothetical prote
15	99.5	6.8	273	1	B64623	lipopolysaccharide
16	99	6.8	492	2	T20368	hypothetical prote
17	95.5	6.5	419	2	T19837	hypothetical prote
18	93.5	6.4	444	2	F71916	probable lipopolys
19	93.5	6.4	1571	2	T14155	zinc finger protei
20	91	6.2	266	2	AG2797	conserved hypothet
21	91	6.2	266	2	H97576	hypothetical prote
22	91	6.2	735	2	S76425	hypothetical prote
23	90.5	6.2	383	2	T30730	hypothetical prote
24	90.5	6.2	404	2	C64597	lipopolysacharide
25	90	6.2	284	1	E64620	lipopolysaccharide
26	89.5	6.1	737	2	C84232	kinase anchor prot
27	89.5	6.1	1378	2	T30173	zinc finger protei
28	88	6.0	292	2	D71894	probable lipopolys
29	88	6.0	694	2	B86559	elongation factor
30	88	6.0	694	2	G72065	translation elonga
31	87.5	6.0	404	2	C71907	probable type II D
32	87.5	6.0	1153	2	A97179	ATP-dependent exon
33	87	6.0	256	2	T00097	hypothetical prote
34	87	6.0	273	2	E71890	probable lipopolys
35	86.5	5.9	354	2	S77119	D-alanine-D-alanin
36	86.5	5.9	845	2	T17291	hypothetical prote
37	86	5.9	258	2	AC2829	conserved hypothet
38	86	5.9	300	2	A97607	hypothetical prote
39	86	5.9	361	2	C83350	probable transcrip
40	86	5.9	5069	2	T17464	rifamycin polyketi
41	84.5	5.8	339	2	B72072	ct389 hypothetical
42	84.5	5.8	339	2	G86551	CT389 hypothetical
43	84.5	5.8	410	2	G81596	conserved hypothet
44	83.5	5.7	402	2	B34735	protein-tyrosine k
45	83.5	5.7	435	2	JN0290	protein-tyrosine k

ALIGNMENTS

RESULT 1
A81971
lacto-N-neotetraose biosynthesis glycosyl transferase NMA0525 [imported] - Neisseria
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: A81971
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo ; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: A81971
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-279 <PAR>
A;Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83817.1; PID:g737
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: lgtB; NMA0525
C;Superfamily: lipopolysaccharide biosynthesis-associated protein

Query Match 91.0%; Score 1329; DB 2; Length 279;
Best Local Similarity 92.5%; Pred. No. 6.1e-115;
Matches 258; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

```
Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||  ||| ||||||||||||||||:|||||||||||||||||
Db      1 MQNHVISLASAAERRAHITDTFGVRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:||||||:|||||||||||||||||:||||| ||||||||||||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYIAVFEDDVLLGEGAEKFLAEDAWLQERFDPDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||||||||||||||| |||||||||| |||||||||||| ||||||| || | |||
Db    121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMWFFLDRFAALPSEGLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          || |||||||||| ||||||||||||||||||||||||||||||||||| || ||| |||
Db    181 VDWMMFGNPDDRERMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRCLNSKQQRRDS 240

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          |||||||||||||||| ||||||||||||||||||||||
Db    241 PANTFKHRLIRALTKISREREKRRQRREQLIGKIIVPFQ 279
```

RESULT 2
C81027
lacto-N-neotetraose biosynthesis glycosyl transferase LgtB NMB1928 [imported] - Neiss
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: C81027

R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: C81027
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-275 <TET>
A;Cross-references: GB:AE002541; GB:AE002098; NID:g7227175; PIDN:AAF42257.1; PID:g722718
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1928
C;Superfamily: lipopolysaccharide biosynthesis-associated protein

Query Match 88.5%; Score 1293; DB 2; Length 275;
Best Local Similarity 90.3%; Pred. No. 1.3e-111;
Matches 252; Conservative 7; Mismatches 16; Indels 4; Gaps 1;

```
Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||||| |||  |||||||||||||||||:||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGRHGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:||||||:||| |||||||||||||:||||| |||||||||:||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYITVFEDDVLLGEGAEKFLAEDAWLQERFDPDTAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||||||||||||||| |||||||||| ||||||||||||||||||| |||| |||
Db    121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMRFFLDRFAALPPEGLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          |||||| :  |||||||||||||||||||||||||||||||||||||||||| |||||| |||
Db    181 VDLMMFSDFFDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRLLNRKQQRRDS 240

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          |||||||||||||||| ||||||||||     |||||
Db    241 PANTFKHRLIRALTKISREREKRRQRREQ----FIVPFQ 275
```

RESULT 3
S70814
glycosyl transferase B (EC 2.4.-.-) - Neisseria meningitidis
C;Species: Neisseria meningitidis
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 20-Sep-1999
C;Accession: S70814
R;Jennings, M.P.; Hood, D.W.; Peak, I.R.A.; Virji, M.; Moxon, E.R.
Mol. Microbiol. 18, 729-740, 1995
A;Title: Molecular analysis of a locus for the biosynthesis and phase-variable expressio
A;Reference number: S70812; MUID:96414473; PMID:8817494
A;Accession: S70814
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-275 <JEN>
A;Cross-references: EMBL:U25839; NID:g973183; PIDN:AAC44085.1; PID:g973186
C;Genetics:
A;Gene: lgtB
C;Superfamily: lipopolysaccharide biosynthesis-associated protein
C;Keywords: glycosyltransferase

Query Match 88.2%; Score 1288; DB 2; Length 275;
Best Local Similarity 90.0%; Pred. No. 3.6e-111;
Matches 251; Conservative 7; Mismatches 17; Indels 4; Gaps 1;

```
Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||||| |||  |||||||||||||||||:||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGRHGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:||||||:||| |||||||||||  |:|||||  |||||||||:||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYITVFEDDVLLGEGEEKFLAEDAWLQERFDPDTAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||||||||||||||| |||||||||| ||||||||||||||||||| |||| |||
Db    121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMRFFLDRFAALPPEGLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          |||||| :  |||||||||||||||||||||||||||||||||||||||||| |||||| |||
Db    181 VDLMMFSDFFDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRLLNRKQQRRDS 240

Qy    241 PANTFKHRLIRALTKISREREKRRQRREQ----FIVPFQ 275
          |||||||||||||||| ||||||||||     |||||
Db    241 PANTFKHRLIRALTKISREREKRRQRREQ----FIVPFQ 275
```

RESULT 4
B81971
lacto-N-neotetraose biosynthesis glycosyl transferase NMA0527 [imported] - Neisseria
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: B81971
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: B81971
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-268 <PAR>
A;Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83819.1; PID:g737
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: lgtA2'; NMA0527
C;Superfamily: lipopolysaccharide biosynthesis-associated protein

Query Match 68.3%; Score 997.5; DB 2; Length 268;
Best Local Similarity 72.9%; Pred. No. 2.3e-84;
Matches 207; Conservative 24; Mismatches 32; Indels 21; Gaps 5;

```
Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||||  ||| ||||||||||||||||:||||||||||||||||||
Db      1 MQNHVISLASAAERRAHITDTFGVRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:||||||:||||||||||||||||||:||||| ||||||||||||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYIAVFEDDVLLGEGAEKFLAEDAWLQERFDPDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||||||||||||||| |||||||||| ||||||||||||||||||| |||| |||
Db    121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMRFFLDRFAALPPEGLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          || |:| :  |: || |||| ||||          |:  | | ::: |: : :   | |
Db    181 VDWMLFSSFLDKGGMTVCQLTPALCV----------QSETLPSQLKNGRQESYRN--RRS 228

Qy    241 PANTFKHRLIRALTKIGRE----REKRRQRR-EQLIGKIIVPFQ 279
          |    |  | ||| |||||    ||::||:: |: :|: :||||:
Db    229 P----KVLLKRALGKIGREIERARERKRQKKLEKHLGRHVVPFE 268
```

RESULT 5
A81027
lacto-N-neotetraose biosynthesis glycosyl transferase LgtE NMB1926 [imported] - Neiss
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: A81027
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307

```
A;Accession: A81027
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-280 <TET>
A;Cross-references: GB:AE002541; GB:AE002098; NID:g7227175; PIDN:AAF42255.1; PID:g722718
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1926
C;Superfamily: lipopolysaccharide biosynthesis-associated protein

  Query Match             63.6%;  Score 928.5;  DB 2;  Length 280;
  Best Local Similarity   65.4%;  Pred. No. 5.6e-78;
  Matches  191;  Conservative   28;  Mismatches   48;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           ||||||||||||||||||||||| |||||||||||||| | | ||||||||:    ||
Db       1 MQNHVISLASAAERRAHIAATFGVRGIPFQFFDALMPSEELNRMMAELVPGLAKQHLLSE 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           |||||||||||||:||||||:||:|||||||||| ||:||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWKQALDEGLPYVAVFEDDVLLGKDAEKFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |: '|  | :|  |:|||||||| |||||||||:||||||:|||||| | :
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHWGTAGYIISREAMRFFLERFAVLPAEWIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           || |||    |:||||| |:||||| ||||||||  :|| ||| :| ||   |
Db     181 VDWMMFTYFFDKEGMPVYQVNPALCTQELHYAKFLSKNSMLGSDLEKDREQER------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
             :||         | ||| | |||::|| :|:     |:  |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKAYGRRVISFK 280


RESULT 6
S70815
glycosyl transferase E (EC 2.4.-.-) - Neisseria meningitidis
C;Species: Neisseria meningitidis
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 20-Sep-1999
C;Accession: S70815
R;Jennings, M.P.; Hood, D.W.; Peak, I.R.A.; Virji, M.; Moxon, E.R.
Mol. Microbiol. 18, 729-740, 1995
A;Title: Molecular analysis of a locus for the biosynthesis and phase-variable expressio
A;Reference number: S70812; MUID:96414473; PMID:8817494
A;Accession: S70815
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-276 <JEN>
A;Cross-references: EMBL:U25839; NID:g973183; PIDN:AAC44086.1; PID:g973187
C;Genetics:
A;Gene: lgtE
C;Superfamily: lipopolysaccharide biosynthesis-associated protein
C;Keywords: glycosyltransferase

  Query Match             60.3%;  Score 881.5;  DB 2;  Length 276;
  Best Local Similarity   63.4%;  Pred. No. 1.2e-73;
  Matches  185;  Conservative   27;  Mismatches   51;  Indels   29;  Gaps    4;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||| ||||||| |||||||||||||| | | ||||||||:    ||
Db       1 MQNHVISLASAAERRGHIAATFGVRGIPFQFFDALMPSEELNRMMAELVPGLAKQHLLSE 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           |||||||||||||:||||||:||:||||    ||: ||:|||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWKQALDEGLPYVAVFE----LGKDAEKFLAEDTWLEERFDKDSAFIVR 116

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:|||||||| |||||||||:||||||:||||||
Db     117 LETMFAKVIVRPDKVLNYENRSFPLLESEHWGTAGYIISREAMRFFLERFAVLPARVDKA 176

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           || |||    |:||||| |:||||| ||||||||  :|| ||| :| ||   |

Db     177 VDWMMFTYFFDKEGMPVYQVNPALCTQELHYAKFLSKNSMLGSDLEKDREQER------- 229

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
             :||         | ||| | |||::|| :|:     |:  |: :: |:
Db     230 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKAYGRRVISFK 276


RESULT 7
S71025
lipopolysaccharide biosynthesis protein lic2B - Haemophilus influenzae (strain RM7004
C;Species: Haemophilus influenzae
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C;Accession: S71025
R;High, N.J.; Jennings, M.P.; Moxon, E.R.
Mol. Microbiol. 20, 165-174, 1996
A;Title: Tandem repeats of the tetramer 5'-CAAT-3' present in lic2A are required for
A;Reference number: S71024; MUID:97014379; PMID:8861214
A;Accession: S71025
A;Molecule type: DNA
A;Residues: 1-266 <HIG>
A;Cross-references: EMBL:Z54182; NID:g1150402; PIDN:CAA90892.1; PID:g1150403
A;Experimental source: strain RM7004
C;Genetics:
A;Gene: lic2B
C;Function:
A;Description: required for the biosynthesis of a phase-variable lipopolysaccharide s
C;Superfamily: lipopolysaccharide biosynthesis-associated protein

  Query Match             31.6%;  Score 461;  DB 2;  Length 266;
  Best Local Similarity   39.1%;  Pred. No. 7.2e-35;
  Matches  109;  Conservative   46;  Mismatches  106;  Indels   18;  Gaps    6;

Qy       3 NHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSGVE 62
           |:|||| || :|| ||   |  : |||:||||| ||: |   : : :| |  |  |:  |
Db       4 NYVISLTSAYQRREHIQKEFSQQNIPFEFFDALKPSKELTSLIEKFIPNL-LHAKLTEGE 62

Qy      63 KACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVRLE 122
           ||||||| :||::   | :||| :||||||||| |::|||| ||:| |     |::| |
Db      63 KACFMSHYMLWQKCFSEDLPYIYIFEDDVLLGENADKFLAEDEWLEEAFKQTDKFILRFE 122

Qy     123 TMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHPVD 182
           |           :  |.||    | ||:|| |||:|||:|::       |    |  :|
Db     123 TFLNFSKCKDKKIKPYSGRKILKLVSENCGAAGYVISREAVKQLSAHICSLTSNHLLAID 182

Qy     183 LMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRR--LNRKQQWRDS 240
           |:||    :       |::| :| ||    : : ::  | | :| :|:  |: |::
Db     183 LLMF----NIFNQSTYQVSPGVCVQE---GQLYPKDIKLHSQLETERQKYLSVKKK---- 231

Qy     241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
             | |  ||    |   ::  |:  |:  | | ||||:
Db     232 --RTLKTVLISLAGK--PKKILRKIYRKLFISKHIVPFR 266


RESULT 8
A64077
lipopolysaccharide biosynthesis protein lic2A - Haemophilus influenzae
C;Species: Haemophilus influenzae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C;Accession: A64077; S15282; S39576
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: A64077
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-302 <TIGR>
A;Cross-references: GB:U32736; GB:L42023; NID:g1573519; PIDN:AAC22208.1; PID:g1573535
A;Experimental source: strain Rd KW20
R;Cope, L.D.; Yogev, R.; Mertsola, J.; Latimer, J.L.; Hanson, M.S.; McCracken Jr., G.
```

```
Mol. Microbiol. 5, 1113-1124, 1991
A;Title: Molecular cloning of a gene involved in lipooligosaccharide biosynthesis and vi
A;Reference number: S15282; MUID:92065807; PMID:1956289
A;Accession: S15282
A;Molecule type: DNA
A;Residues: 1-28,'N',30-40,45-105,'L',107-150,'R',152-302 <COP>
A;Cross-references: EMBL:X56903; NID:g48811; PIDN:CAA40221.1; PID:g48812
A;Experimental source: strain DL42
A;Note: it is uncertain whether Met-1 or Met-11 is the initiator
R;High, N.J.; Deadman, M.E.; Moxon, E.R.
Mol. Microbiol. 9, 1275-1282, 1993
A;Title: The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of t
A;Reference number: S39576; MUID:95020659; PMID:7523834
A;Accession: S39576
A;Molecule type: DNA
A;Residues: 1-28,'N',30-40,49-105,'L',107-150,'R',152-255,'P',257-273,'D',275-302 <HIG>
A;Cross-references: EMBL:L19441; NID:g305379; PIDN:AAA65534.1; PID:g305380
A;Experimental source: strain RM7004
C;Genetics:
A;Gene: lic2A
C;Function:
A;Description: involved in lipooligosaccharide biosynthesis and virulence expression
C;Superfamily: lipopolysaccharide biosynthesis-associated protein

  Query Match             28.6%;  Score 418;  DB 1;  Length 302;
  Best Local Similarity   32.9%;  Pred. No. 7.9e-31;
  Matches  100;  Conservative   63;  Mismatches  109;  Indels   32;  Gaps    6;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDAL------------------MPSERL 41
           ::| |||: :| ||| ||   | |: : | ||:|                  :: :
Db       4 IENIVISMENATERRKHITKQFESKKLSFSFFNAYTYQSINQSINQSINQSINQSI 63

Qy      42 ERAMAE---LVPGLSAHPYLSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAE 98
             ::: :   ::  :     |:  || | :|| :|| : ::|   |: :|||||:||| ||
Db      64 NQSINQSNSILHNIEESRILTKGEKGCLISHFLLWNKCVNENFEYLKIFEDDVILGENAE 123

Qy      99 QFLAEDTWLQERFDPDSAFVVRLETMFMHV-LTSPSGVADYGGRAFPLLESEHCGTAGYI 157
           || ::.||: ||| :  |::||||    | |   : :  :  | | :|:| | ||||||
Db-    124 VFLNQNEWLKTRFDFNDIFIIRLETFLQPVKLEKQTKIPPFNSRNFDILKSTHWGTAGYI 183

Qy     158 ISRKAMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQ 217
           ||: | :: ::    :| : :  || ::|    | :   | |||||:| |||   : :
Db     184 ISQGAAKYVIEYLKNIPSDEIVAVDELIFNKLVDVDNYIVYQLNPAICIQEL---QANQS 240

Qy     218 NSALGSLIEHDRRLN---RKQQWRDSPANTFKHRLIRALTKIGREREKRRQRREQLIGKI 274
           | | | :| :|:     ||::          |  :||||    |:: | :|| ::: ||
Db     241 KSVLTSGLEKERQKRSKIRKKKTLKQRLTRIKENIIRAL---NRKKWKEQQRIKEMQGKE 297

Qy     275 IVPF 278
           || |
Db     298 IVRF 301


RESULT 9
F64091
lipopolysaccharide biosynthesis protein lic2B homolog - Haemophilus influenzae (strain R
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 20-Sep-1999
C;Accession: F64091
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: F64091
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-282 <TIGR>
A;Cross-references: GB:U32760; GB:L42023; NID:g1573764; PIDN:AAC22423.1; PID:g1573773; T
C;Genetics:
A;Start codon: GTG
C;Superfamily: lipopolysaccharide biosynthesis-associated protein

  Query Match             18.5%;  Score 271;  DB 2;  Length 282;
  Best Local Similarity   29.0%;  Pred. No. 2.6e-17;
  Matches   80;  Conservative   43;  Mismatches  103;  Indels   50;  Gaps    9;

Qy       3 NHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHP--YLSG 60
           |:|||| :  :|| ||   || : |||:||||: | : :|    :    |   |   ||
Db      34 NYVISLTTEQKRRKHITEEFGKQNIPFEFFDAITP-DIIEETAKKFNITLDRSPKAKLSD 92

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
             |  | :|| |||: ||:  : || :||||: ||| |:: |  |      :  |   |::
Db      93 GEIGCALSHIVLWDLALENNLNYINIFEDDIHLGENAKELLEID------YISDDIHVLK 146

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           ||          |  |        :|:   :  | ||| :: |  :: |:     | :
Db     147 LEANGKMFFKQPKSV-KCDRNVYPMTVKQ-SGCAGYTVTAKGAKYLLELVKNKPLD--VA 202

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQE--LHYAKFHDQNSALGSLIEHDRRLNRKQQWR 238
           || ::| :    :   : ||:| :| |:  ||
Db     203 VDSLVFEDFLHFKDYKIVQLSPGICVQDFVLH---------------------------- 234

Qy     239 DSPANTFKHRLIRALTKIGREREKRRQRREQLIGKI 274
             | |:  |     : ||:|    ||:  :: ||
Db     235 --PDNPFESSL-----QEGRDRVHGNQRKSSILEKI 263


RESULT 10
AD3304
lipooligosaccharide biosynthesis protein lic2B [imported] - Brucella melitensis (stra
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AD3304
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A;Reference number: AD3252; PMID:11756688
A;Accession: AD3304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-268 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL51599.1; PID:g17982323; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0418
A;Map position: I

  Query Match             15.8%;  Score 231.5;  DB 2;  Length 268;
  Best Local Similarity   28.6%;  Pred. No. 1.1e-13;
  Matches   82;  Conservative   43;  Mismatches   95;  Indels   67;  Gaps   12;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLE----RAMAEL-------- 48
           |: ::|:|  : :|   :|: |   |  |          ||:|    |||: |
Db       9 MKCYLINLDKSRDRLEFMASQFERLGAQF---------ERVEAVNGRAMSPLELASFTQI 59

Qy      49 -----VPGLSAHPYLSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAE 103
                |       ||  |  ||:||    |:        | ||||||: | :|: :|||
Db      60 SKEWPAP-------LSPAEIGCFLSHRKCLEKIAAGEDAYAAVFEDDIRLSQGSSRFLAS 112

Qy     104 DTWLQERFDPDSAFVVRLETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAM 163
           | |:     |  | :|:::     || | : | : |  :   | | |  | ||::||:|
Db     113 DHWI-----PKQADIVKIDAYGHEVLIS-NPVKNEGPYSISRLRSRHLQTGGYVVSREAA 166

Qy     164 RFFLDRFAVLPPERLHPVDLMMFGNPDDR--EGMPVCQLNPALCAQELHYAKFHDQNSAL 221
           |  |     |  :   |||  :| :|:|       : |::||:| |           | :
Db     167 RKLLP----LMEKVSAPVDHFLF-DPNDGPFNDFEIYQISPAICRQ----------SGM 210

Qy     222 GSLIEHDRRLNRKQQ-----WRDSPANTFKHR-----LIRALTKIGR 258
           | |  :||  ::       ||::     : |       |  :|| ||
Db     211 ESTIGQNRRPKQRPSLLGLVWREAKRLVMRTRRNLKGFIANVTKTGR 257
```

```
RESULT 11
E71916
probable lipopolysaccharide biosynthesis protein - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C;Accession: E71916
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: E71916
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-332 <ARN>
A;Cross-references: GB:AE001489; GB:AE001439; NID:g4155102; PIDN:AAD06135.1; PID:g415510
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp0562

  Query Match             8.8%;  Score 128;  DB 2;  Length 332;
  Best Local Similarity   23.8%;  Pred. No. 0.00051;
  Matches   77;  Conservative   36;  Mismatches  108;  Indels  102;  Gaps   17;

Qy      29 FQFFDALMPSER-LERAMAELVPGLSA------HPY----LSGVEKACFMSHAVLWEQAL 77
           || |||: |  :  |: : ||    |       | |    |:  |  |:::|| :||:: :
Db      35 FQIFDAISPKHQDFEKLLQELYNAQSLLQSDWYHSYVGAGLTLPELGCYLSHYLLWKECV 94

Qy      78 DEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVRL--------ETMFMHVL 129
             | : : |||| |     | | ||  |:  ||      |||         || | |||
Db      95 KLDQP-VVILEDDVTLESHFMQAL-EDC-LKSPFD-----FVRLYGCYWYYHETKF-HVL 145

Qy     130 TSPSGVA----DYGGRAFPLLES---------------------------------- 148
            |       |:  :  |:||
Db     146 --PKEFVFPPFDHSFKNNPILEKFKKFFDVSRFLNLSTHKVIHYILKKIQKSYYATHEKE 203

Qy     149 ----EH--------CGTAGYIISRKAMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMP 196
               ||          |||| :: |  : |::  |    : : |||: |    |:
Db     204 AFFLEHFYLTSVYVASTAGYYLTPKGAKTFIE--ATESFKIIEPVDMFM----DNSAYHD 257

Qy     197 VCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDSPANTFKHRLIRALTKI 256
           |  |    |   |          :| | |:  :: : |        :|||:    :|
Db     258 VANLTYVPCPVSL-------SEHSLDSTIQKPQKKSLKSYPLPPQKSTFKNLFYYSL--- 307

Qy     257 GREREKRRQRREQLIGKIIVPFQ 279
             |:|    |   |   | :
Db     308 ---NAKKRLNAFQQYSKQYAPLK 327


RESULT 12
AH3417
lpsA protein [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AH3417
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AH3417
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-721 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL52507.1; PID:g17983318; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI1326
A;Map position: I

  Query Match             8.0%;  Score 116.5;  DB 2;  Length 721;
  Best Local Similarity   25.8%;  Pred. No. 0.016;
  Matches   48;  Conservative   27;  Mismatches   74;  Indels   37;  Gaps    7;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQF--FDALMPSERLERA----------MAELVPG 51
           ||:::||       :||:  : |  ||    ||:      :|             |::||
Db      25 VINMASQPAAYKTVAASIEAYGQGFQLHRIDAVNGHTATQRIGIDDARFDAINGREMLPG 84

Qy      52 LSAHPYLSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERF 111
                 |  |: ||:   |  | :| ||  : ||||:  |      : :       |
Db      85 ----------EYGCYRSHSKALESFLSDGSPYGLILEDDVVFTETTSARIHDIIKSLPDF 134

Qy     112 DPDSAFVVRL----ETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFL 167
           |     ||:|      :|| :| : :|  |  |||   :       | |::||:  |  |
Db     135 D-----VVKLVNHRSPLFMSLLETDAG--DRIGRA---IHGPQGSAAAYLVSREGARKLL 184

Qy     168 DRFAVL 173
               : :
Db     185 SALSTM 190


RESULT 13
T00343
hypothetical protein KIAA0584 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00343
R;Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
DNA Res. 5, 31-39, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A;Reference number: Z14086; MUID:98290545; PMID:9628581
A;Accession: T00343
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-738 <NAG>
A;Cross-references: EMBL:AB011156; NID:g3043691; PIDN:BAA25510.1; PID:g3043692
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0584

  Query Match             7.3%;  Score 107;  DB 2;  Length 738;
  Best Local Similarity   23.3%;  Pred. No. 0.13;
  Matches   44;  Conservative   40;  Mismatches   83;  Indels   22;  Gaps    7;

Qy       5 VISLASAAERRAHIAATFGSRGIPFQFFDAL----MPSERLERAMAELVPG----LSAHP 56
           :|:|    :||  :  |   : |  :  :|:    : : :|:    |::||     |: |
Db     458 MINLKRRKDRRDRMLRTLYEQEIEVKIVEAVDGKALNTSQLKALNIEMLPGYRDPYSSRP 517

Qy      57 YLSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAE--DTWLQERFDPD 114
             | |  ||:||  :|:: :|  :    | ||||      :: | :  |   | : | :
Db     518 LTRG-EIGCFLSHYSVWKEVIDRELEKTLVIEDDVRFEHQFKKKLMKLMDNIDQAQLDWE 576

Qy     115 SAFVVRLETMFMHVLTSPSGVADYGGRAFPLLESEHC-GTAGYIISRKAMRFFLDRFAVL 173
           :: |          :   ||:       |:|:::   | ||:|| :  :  :
Db     577 LIYIGRKRMQVKEPEKAVPNVAN-------LVEADYSYWTLGYVISLEGAQKLV---GAN 626

Qy     174 PPERLHPVD 182
           |  :: |||
Db     627 PFGKMLPVD 635


RESULT 14
E64175
hypothetical protein HI1697 (lsg locus) - Haemophilus influenzae
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C;Accession: E64175; S27580
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: E64175
```

```
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-267 <TIGR>
A;Cross-references: GB:L42023; TIGR:HI1697
A;Experimental source: strain Rd KW20
R;McLaughlin, R.; Abu Kwaik, Y.; Young, R.; Spinola, S.; Apicella, M.
submitted to the EMBL Data Library, June 1992
A;Description: Characterization and sequence of the lsg locus from Haemophilus influenza
A;Reference number: S27577
A;Accession: S27580
A;Molecule type: DNA
A;Residues: 11-26,'K',28-39,'I',41-55,'S',57-111,'L',113-136,'N',138-139,'N',141-143,'T'
A;Cross-references: EMBL:M94855; NID:g148931; PIDN:AAA24981.1; PID:g148935
A;Experimental source: strain A2

  Query Match             6.9%;  Score 100.5;  DB 2;  Length 267;
  Best Local Similarity   23.0%;  Pred. No. 0.13;
  Matches   65;  Conservative   42;  Mismatches  113;  Indels   63;  Gaps   15;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGI-PFQFFDAL--MPSERLERAMAELVPGLSAHPY 57
           :: ::|||    :||      |  :    || | |:  |  :  | |    :    || |
Db      12 LKKYLISLDKDIQRR---ELFFSQKNTEDFQVFSAINTMQKDWDELAAIFNIEQFKAH-Y 67

Qy      58 LSGVEK---ACFMSHAVLWEQALDEG----VPYIAVFEDDVLLGEGAEQFLAEDTWLQER 110
             | |    | :||  :::: :::       |  | ||| |     :| |     | |:
Db      68 FRNVTKGEIGCTLSHLSVYQKIVEDNDIAEDSYALVCEDDALFHSDFQQNLT--ALLSEK 125

Qy     111 FDP--------------DSAFVVRLETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGY 156
            :               |    :   | |   |   :|  :|   |||  :|   || ||
Db     126 LEAEIILLGQSKINDFNDFDLEINYPTTF-SFLCKKTGDVNY---AFP-YKSYFAGTVGY 180

Qy     157 IISRKAMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHD 216
           :| : | | |: : :   |  |   | ::|                          |:
Db     181 LIKKSAARRFIQQISQNKPFWLAD-DFLLFEQ-----------------------NFNI 215

Qy     217 QNSALGSL--IEHDRRLNRKQQWRDSPANTFKHRLIR-ALTKI 256
           :|  :  |  ||:   ::  :  | | :|    :|::  | ||
Db     216 RNKVVRPLIVIENPVLISNLESVRGSLSNNLLKKLMKYPLKKI 258


RESULT 15
B64623
lipopolysaccharide biosynthesis-associated protein lex2B - Helicobacter pylori (strain 2
C;Species: Helicobacter pylori
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jun-2000
C;Accession: B64623
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: B64623
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-273 <TOM>
A;Cross-references: GB:AE000594; GB:AE000511; NID:g2313957; PIDN:AAD07876.1; PID:g231396
C;Genetics:
A;Start codon: TTG
C;Superfamily: lipopolysaccharide biosynthesis-associated protein

  Query Match             6.8%;  Score 99.5;  DB 1;  Length 273;
  Best Local Similarity   23.7%;  Pred. No. 0.17;
  Matches   58;  Conservative   32;  Mismatches   74;  Indels   81;  Gaps   15;

Qy      30 QFFDALMPSERLERAMAELVPGLSAHPY--------------LSGVEK---------- 63
           | |||:  |:  |  :  ||     |||               :| | |
Db      38 QIFDAIY-SKTFEGGLHPLVK-KHLHPYFITQNIKDMGITTNLISEVSKFYYALKYHAKF 95

Qy      64 ------ACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAF 117
                 |: ||  |||: ::     | : |||: | |  :: |    :|::
Db      96 MSLGELGCYASHYSLWEKCIELN-EAICILEDDITLKEDFKEGL---DFLEKHI--QELG 149

Qy     118 VVRLETMFMHVLTSPSGVADYGGRAFPLLESEH--------------CGTAGYIISRKA 162
             :||    ||:|       |   :: ||    |                || ||:|: |
Db     150 YIRL----MHLL------YDASVKSEPLSHKNHEIQERVGIIKAYSEGVGTQGYVITPKI 199

Qy     163 MRFFL--DRFAVLPPER------LHPVD---LMMFGNPDDREGMPVCQ----LNPALC-A 206
             : ||   |  |:| :       :| |    |  |   || :   : :     :| :
Db     200 AKVFLKCSRKWVVPVDTIMDATFIHGVKNLVLQPFVIADDEQISTIARKEEPYSPKIALM 259

Qy     207 QELHY 211
           :|||:
Db     260 RELHF 264


Search completed: May 27, 2003, 18:18:10
Job time : 47 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 27, 2003, 18:06:19 ; Search time 11 Seconds
(without alignments)
1051.991 Million cell updates/sec

Title: US-10-007-267A-8
Perfect score: 1461
Sequence: 1 MQNHVISLASAAERRAHIAA.....REKRRQRREQLIGKIIVPFQ 279

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1461	100.0	279	1	LGTB_NEIGO	Q50947 neisseria g
2	1329	91.0	279	1	LGTB_NEIMA	P57033 neisseria m
3	1293	88.5	275	1	LGTB_NEIMB	Q51116 neisseria m
4	1036.5	70.9	280	1	LGTE_NEIGO	Q50950 neisseria g
5	928.5	63.6	280	1	LGTE_NEIMB	Q51117 neisseria m
6	461	31.6	266	1	LI2B_HAEIN	Q57394 haemophilus
7	418	28.6	302	1	LEX1_HAEIN	Q03974 haemophilus
8	321	22.0	263	1	LPSA_PASHA	Q05770 pasteurella
9	271	18.5	282	1	Y765_HAEIN	Q57125 haemophilus
10	100.5	6.9	257	1	LSG4_HAEIN	P71398 haemophilus
11	88	6.0	694	1	EFG_CHLPN	Q9z802 chlamydia p
12	86.5	5.9	354	1	DDL_SYNY3	P73632 synechocyst
13	83.5	5.7	741	1	PIFA_ECOLI	P96329 escherichia
14	83.5	5.7	971	1	KBF1_MOUSE	P25799 mus musculu
15	83.5	5.7	1114	1	RET_HUMAN	P07949 homo sapien
16	82.5	5.6	273	1	NADC_HELPY	O25909 helicobacte
17	82	5.6	1045	1	RAG1_XENLA	Q91829 xenopus lae
18	81	5.5	275	1	HUPH_RHOCA	Q03007 rhodobacter
19	81	5.5	380	1	CAIA_SALTY	Q8zrx2 salmonella
20	81	5.5	554	1	Y514_SYNY3	Q55470 synechocyst
21	81	5.5	726	1	RNR_MYCPN	P75529 mycoplasma
22	81	5.5	790	1	ATSY_SYNP7	P37385 synechococc
23	80.5	5.5	429	1	MYCS_RAT	P23999 rattus norv
24	80	5.5	1237	1	B3A2_RABIT	P48746 oryctolagus
25	79.5	5.4	462	1	TRPE_LEPBI	P20463 leptospira
26	79.5	5.4	968	1	KBF1_HUMAN	P19838 homo sapien
27	79	5.4	522	1	KBF1_RAT	Q63369 rattus norv
28	79	5.4	933	1	GLND_RHILO	Q98c27 rhizobium l
29	78.5	5.4	344	1	KAND_SCHPO	Q10078 schizosacch
30	78.5	5.4	1268	1	YRD6_CAEEL	Q09575 caenorhabdi
31	78	5.3	359	1	Y130_ARCFU	O30107 archaeoglob
32	78	5.3	395	1	LDOX_MAIZE	P41213 zea mays (m
33	78	5.3	424	1	CRT3_ARATH	O04153 arabidopsis
34	78	5.3	1234	1	B3A2_RAT	P23347 rattus norv
35	78	5.3	1237	1	B3A2_MOUSE	P13808 mus musculu
36	78	5.3	1238	1	B3A2_CAVPO	Q9z0s8 cavia porce
37	78	5.3	1241	1	B3A2_HUMAN	P04920 homo sapien
38	77.5	5.3	273	1	NADC_HELPJ	Q9zjn2 helicobacte
39	77.5	5.3	380	1	CAIA_SALTI	Q8z912 salmonella
40	77.5	5.3	965	1	CAPP_SOLTU	P29196 solanum tub
41	77	5.3	694	1	EFG_CHLTR	O84444 chlamydia t
42	76.5	5.2	694	1	EFG_CHLMU	Q9pjv6 chlamydia m
43	76.5	5.2	967	1	CAPP_PEA	P51062 pisum sativ
44	76.5	5.2	1115	1	RET_MOUSE	P35546 mus musculu
45	76	5.2	949	1	GLND_RHIME	P56884 rhizobium m

ALIGNMENTS

RESULT 1
LGTB_NEIGO
ID LGTB_NEIGO STANDARD; PRT; 279 AA.
AC Q50947;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lacto-N-neotetraose biosynthesis glycosyl transferase lgtB.
GN LGTB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F62;
RX MEDLINE=95053752; PubMed=7964493;
RA Gotschlich E.C.;
RT "Genetic locus for the biosynthesis of the variable portion of
RT Neisseria gonorrhoeae lipooligosaccharide.";
RL J. Exp. Med. 180:2181-2190(1994).
CC -!- FUNCTION: ADDS THE SECOND GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
CC IN LOS.
CC -!- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
CC STRUCTURE.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC ----------

CC ----------
DR EMBL; U14554; AAA68010.1; -.
DR InterPro; IPR002654; GT_25.
DR Pfam; PF01755; Glyco_transf_25; 1.
KW Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.
SQ SEQUENCE 279 AA; 31776 MW; C2219D1A6119D622 CRC64;

Query Match 100.0%; Score 1461; DB 1; Length 279;
Best Local Similarity 100.0%; Pred. No. 6.9e-128;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
```

```
Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          |||||||||||||||||||||||||||||||||||||||
Db    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279


RESULT 2
LGTB_NEIMA
ID   LGTB_NEIMA     STANDARD;      PRT;   279 AA.
AC   P57033;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lacto-N-neotetraose biosynthesis glycosyl transferase lgtB.
GN   LGTB OR NMA0525.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
CC   -!- FUNCTION: ADDS THE SECOND GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
CC       IN LOS.
CC   -!- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
CC       STRUCTURE.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL162753; CAB83817.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;
KW   Complete proteome.
SQ   SEQUENCE   279 AA;  31903 MW;  8703B56513A0D347 CRC64;

  Query Match             91.0%;  Score 1329;  DB 1;  Length 279;
  Best Local Similarity   92.5%;  Pred. No. 1.1e-115;
  Matches  258;  Conservative    5;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||  ||| ||||||||||||||||:||||||||||||||||||
Db      1 MQNHVISLASAAERRAHITDTFGVRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:||||||:||||||||||||||||||:||||| ||||||||||||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYIAVFEDDVLLGEGAEKFLAEDAWLQERFDPDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||||||||||||||| |||||||||| |||||||||||| ||||||| || | |||
Db    121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMWFFLDRFAALPSEGLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          || |||||||||| |||||||||||||||||||||||||||||||||| || ||| |||
Db    181 VDWMMFGNPDDRERMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRCLNSKQQRRDS 240

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          |||||||||||||| ||||||||||||||||||||||||
Db    241 PANTFKHRLIRALTKISREREKRRQRREQLIGKIIVPFQ 279


RESULT 3
LGTB_NEIMB
ID   LGTB_NEIMB     STANDARD;      PRT;   275 AA.
AC   Q51116;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lacto-N-neotetraose biosynthesis glycosyl transferase lgtB.
GN   LGTB OR NMB1928.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / Serogroup B;
RX   MEDLINE=96414473; PubMed=8817494;
RA   Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;
RT   "Molecular analysis of a locus for the biosynthesis and phase-variable
RT   expression of the lacto-N-neotetraose terminal lipopolysaccharide
RT   structure in Neisseria meningitidis.";
RL   Mol. Microbiol. 18:729-740(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / Serogroup B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
CC   -!- FUNCTION: ADDS THE SECOND GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
CC       IN LOS.
CC   -!- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
CC       STRUCTURE.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U25839; AAC44085.1; -.
DR   EMBL; AE002541; AAF42257.1; -.
DR   TIGR; NMB1928; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;
KW   Complete proteome.
FT   CONFLICT     97      97       A -> E (IN REF. 1).
SQ   SEQUENCE   275 AA;  31578 MW;  E871305E2F6CF70F CRC64;

  Query Match             88.5%;  Score 1293;  DB 1;  Length 275;
  Best Local Similarity   90.3%;  Pred. No. 2.4e-112;
  Matches  252;  Conservative    7;  Mismatches   16;  Indels    4;  Gaps    1;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||| |||  |||||||||||||||||:||||||||||||||||||
```

```
Db      1 MQNHVISLASAAERRAHIADTFGRHGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||:||||||:||| ||||||||||||:||||| ||||||||||:||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYITVFEDDVLLGEGAEKFLAEDAWLQERFDPDTAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||||||||||||| |||||||||| |||||||||||||||||| |||| |||
Db    121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMRFFLDRFAALPPEGLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          |||||| :  |||||||||||||||||||||||||||||||||||||| |||||| |||
Db    181 VDLMMFSDFFDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRLLNRKQQRRDS 240

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          ||||||||||||||| |||||||||||     |||||
Db    241 PANTFKHRLIRALTKISREREKRRQRREQ----FIVPFQ 275


RESULT 4
LGTE_NEIGO
ID   LGTE_NEIGO     STANDARD;      PRT;   280 AA.
AC   Q50950;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lacto-N-neotetraose biosynthesis glycosyl transferase lgtE.
GN   LGTE.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F62;
RX   MEDLINE=95053752; PubMed=7964493;
RA   Gotschlich E.C.;
RT   "Genetic locus for the biosynthesis of the variable portion of
RT   Neisseria gonorrhoeae lipooligosaccharide.";
RL   J. Exp. Med. 180:2181-2190(1994).
CC   -!- FUNCTION: ADDS THE FIRST GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
CC       IN LOS.
CC   -!- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
CC       STRUCTURE.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U14554; AAA68013.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.
SQ   SEQUENCE   280 AA;  32420 MW;  F043597BA3040407 CRC64;

  Query Match             70.9%;  Score 1036.5;  DB 1;  Length 280;
  Best Local Similarity   71.6%;  Pred. No. 1.3e-88;
  Matches  209;  Conservative   24;  Mismatches   34;  Indels   25;  Gaps   3;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||| |||||||||||||||||:||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||||||||:|||||||||||||||||||||||:|||| ||||:||
Db     61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||  |:  |  | :|  |:|||||||||||||||||:|||||||||||||||:
Db    121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||    |:||||| |::|||| ||||||||  ||| ||| :| ||   |
Db    181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy    241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:: |: :: |:
Db    234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 5
LGTE_NEIMB
ID   LGTE_NEIMB     STANDARD;      PRT;   280 AA.
AC   Q51117;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lacto-N-neotetraose biosynthesis glycosyl transferase lgtE.
GN   LGTE OR NMB1926.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / Serogroup B;
RX   MEDLINE=96414473; PubMed=8817494;
RA   Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;
RT   "Molecular analysis of a locus for the biosynthesis and phase-variable
RT   expression of the lacto-N-neotetraose terminal lipopolysaccharide
RT   structure in Neisseria meningitidis.";
RL   Mol. Microbiol. 18:729-740(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / Serogroup B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
CC   -!- FUNCTION: ADDS THE FIRST GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
CC       IN LOS.
CC   -!- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
CC       STRUCTURE.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U25839; AAC44086.1; -.
DR   EMBL; AE002541; AAF42255.1; -.
DR   TIGR; NMB1926; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;
KW   Complete proteome.
FT   CONFLICT     16     16       A -> G (IN REF. 1).
FT   CONFLICT     89     92       MISSING (IN REF. 1).
FT   CONFLICT    176    178       EWI -> RVD (IN REF. 1).
SQ   SEQUENCE   280 AA;  32790 MW;  826827E942BA5842 CRC64;
```

```
  Query Match           63.6%;  Score 928.5;  DB 1;  Length 280;
  Best Local Similarity  65.4%;  Pred. No. 1.3e-78;
  Matches  191;  Conservative   28;  Mismatches   48;  Indels   25;  Gaps    3;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||||||||| |||||||||||||| | | ||||||||:    ||
Db      1 MQNHVISLASAAERRAHIAATFGVRGIPFQFFDALMPSEELNRMMAELVPGLAKQHLLSE 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:||||||:||:|||||||||: ||:|||||||:|||| ||||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYVAVFEDDVLLGKDAEKFLAEDTWLEERFDKDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||  |:  |  | :|  |:||||||| |||||||||:||||||:|||||| | :
Db    121 LETMFAKVIVRPDKVLNYENRSFPLLESEHWGTAGYIISREAMRFFLERFAVLPAEWIKA 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          || |||    |:|||||:|||||| ||||||||  :|| ||| :| ||   |
Db    181 VDWMMFTYFFDKEGMPVYQVNPALCTQELHYAKFLSKNSMLGSDLEKDREQER------- 233

Qy    241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
               :||        | ||| | |||::|| :|:     |:  |: :: |:
Db    234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKAYGRRVISFK 280


RESULT 6
LI2B_HAEIN
ID   LI2B_HAEIN     STANDARD;      PRT;   266 AA.
AC   Q57394;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lipooligosaccharide biosynthesis protein lic2B.
GN   LIC2B.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RM 7004 / Serotype B;
RA   High N.J., Jennings M.P., Moxon R.E.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN EXTRACELLULAR LIPOOLIGOSACCHARIDES (LOS)
CC       BIOSYNTHESIS AND VIRULENCE EXPRESSION. INVOLVED IN THE SYNTHESIS
CC       OF THE OLIGOSACCHARIDE MOIETY OF THE LOS MOLECULE BY ADDING GALAC
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z54182; CAA90892.1; -.
DR   EMBL; U36398; AAA84947.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase; Glycosyltransferase; Virulence.
SQ   SEQUENCE   266 AA;  31131 MW;  0C68AE9BA249BD39 CRC64;

  Query Match           31.6%;  Score 461;  DB 1;  Length 266;
  Best Local Similarity  39.1%;  Pred. No. 2.1e-35;
  Matches  109;  Conservative   46;  Mismatches  106;  Indels   18;  Gaps    6;

Qy      3 NHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSGVE 62
          |:|||| || :|| ||   |  : |||:||||| ||: |   : : :| |  |  |:  |
Db      4 NYVISLTSAYQRREHIQKEFSQQNIPFEFFDALKPSKELTSLIEKFIPNL-LHAKLTEGE 62

Qy     63 KACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVRLE 122
          ||||||| :||::   | :||| :||||||||| |::||||| ||:| |     |::| |
Db     63 KACFMSHYMLWQKCFSEDLPYIYIFEDDVLLGENADKFLAEDEWLEEAFKQTDKFILRFE 122

Qy    123 TMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHPVD 182
          |           :  | ||    | ||:|| |||:|||:|::        |    |  :|
Db    123 TFLNFSKCKDKKIKPYSGRKILKLVSENCGAAGYVISREAVKQLSAHICSLTSNHLLAID 182

Qy    183 LMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRR--LNRKQQWRDS 240
          |:||    :       |::| :| ||    : : ::  | | :| :|:  |: |::
Db    183 LLMF----NIFNQSTYQVSPGVCVQE---GQLYPKDIKLHSQLETERQKYLSVKKK---- 231

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
               | |  ||    |   ::  |:  |:  | | ||||:
Db    232 --RTLKTVLISLAGK--PKKILRKIYRKLFISKHIVPFR 266


RESULT 7
LEX1_HAEIN
ID   LEX1_HAEIN     STANDARD;      PRT;   302 AA.
AC   Q03974; Q48209;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lipooligosaccharide biosynthesis protein lex-1.
GN   LEX1 OR LIC2A OR HI0550.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DL42 / Serotype B;
RX   MEDLINE=92065807; PubMed=1956289;
RA   Cope L.D., Yogev R., Mertsola J., Latimer J.L., Hanson M.S.,
RA   McCracken G.H. Jr., Hansen E.J.;
RT   "Molecular cloning of a gene involved in lipooligosaccharide
RT   biosynthesis and virulence expression by Haemophilus influenzae type
RT   B.";
RL   Mol. Microbiol. 5:1113-1124(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RM 7004 / Serotype B;
RX   MEDLINE=95020659; PubMed=7523834;
RA   High N.J., Deadman M.E., Moxon E.R.;
RT   "The role of a repetitive DNA motif (5'-CAAT-3') in the variable
RT   expression of the Haemophilus influenzae lipopolysaccharide epitope
RT   alpha Gal(1-4)beta Gal.";
RL   Mol. Microbiol. 9:1275-1282(1993).
CC   -!- FUNCTION: INVOLVED IN EXTRACELLULAR LIPOOLIGOSACCHARIDES (LOS)
CC       BIOSYNTHESIS AND VIRULENCE EXPRESSION. INVOLVED IN THE SYNTHESIS
CC       OF THE OLIGOSACCHARIDE MOIETY OF THE LOS MOLECULE BY ADDING GALAC.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-11 IS THE INITIATOR.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32736; AAC22208.1; -.
DR   EMBL; X56903; CAA40221.1; -.
DR   EMBL; L19441; AAA65534.1; -.
DR   PIR; S15282; S15282.
DR   TIGR; HI0550; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase; Glycosyltransferase; Virulence; Complete proteome.
FT   VARIANT      29     29       K -> N (IN STRAINS DL42 AND RM 7004).
FT   VARIANT      63     70       MISSING (IN STRAIN RM 7004).
FT   VARIANT      67     70       MISSING (IN STRAIN DL42).
FT   VARIANT     106    106       F -> L (IN STRAINS DL42 AND RM 7004).
FT   VARIANT     151    151       Q -> R (IN STRAINS DL42 AND RM 7004).
FT   VARIANT     256    256       S -> P (IN STRAIN RM 7004).
FT   VARIANT     274    274       N -> D (IN STRAIN RM 7004).
SQ   SEQUENCE   302 AA;  35490 MW;  BFC204F9B3372D2C CRC64;

  Query Match             28.6%;  Score 418;  DB 1;  Length 302;
  Best Local Similarity   32.9%;  Pred. No. 2.4e-31;
  Matches  100;  Conservative   63;  Mismatches  109;  Indels   32;  Gaps    6;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDAL------------------MPSERL 41
           ::| |||: :| ||| ||   | |: : | ||:|                    :: :
Db       4 IENIVISMENATERRKHITKQFESKKLSFSFFNAYTYQSINQSINQSINQSINQSINQSI 63

Qy      42 ERAMAE---LVPGLSAHPYLSGVEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAE 98
             ::: :   ::  :     |:  || | :|| :|| : ::|   |: :|||||:||| ||
Db      64 NQSINQSNSILHNIEESRILTKGEKGCLISHFLLWNKCVNENFEYLKIFEDDVILGENAE 123

Qy      99 QFLAEDTWLQERFDPDSAFVVRLETMFMHV-LTSPSGVADYGGRAFPLLESEHCGTAGYI 157
            || :: ||: ||| :  |::||||    | |   : :  :  | | :|:| | ||||||
Db     124 VFLNQNEWLKTRFDFNDIFIIRLETFLQPVKLEKQTKIPPFNSRNFDILKSTHWGTAGYI 183

Qy     158 ISRKAMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQ 217
           ||: | :: ::    :| : :  || ::|    | :   | |||||:| |||   : :
Db     184 ISQGAAKYVIEYLKNIPSDEIVAVDELIFNKLVDVDNYIVYQLNPAICIQEL---QANQS 240

Qy     218 NSALGSLIEHDRRLN---RKQQWRDSPANTFKHRLIRALTKIGREREKRRQRREQLIGKI 274
           | | | :| :|:     ||::          |  :||||    |:: | :|| ::: ||
Db     241 KSVLTSGLEKERQKRSKIRKKKTLKQRLTRIKENIIRAL---NRKKWKEQQRIKEMQGKE 297

Qy     275 IVPF 278
           || |
Db     298 IVRF 301


RESULT 8
LPSA_PASHA
ID   LPSA_PASHA     STANDARD;      PRT;   263 AA.
AC   Q05770;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   LpsA protein.
GN   LPSA.
OS   Pasteurella haemolytica.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Mannheimia.
OX   NCBI_TaxID=75985;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Serotype A1;
RX   MEDLINE=95301198; PubMed=7781993;
RA   Potter M.D., Lo R.Y.C.;
RT   "Cloning and characterization of a gene from Pasteurella haemolytica
RT   A1 involved in lipopolysaccharide biosynthesis.";
RL   FEMS Microbiol. Lett. 129:75-81(1995).
RN   [2]
RP   SEQUENCE OF 189-263 FROM N.A.
RC   STRAIN=Serotype A1;
RX   MEDLINE=91358346; PubMed=1885539;
RA   Abdullah K.M., Lo R.Y.C., Mellors A.;
RT   "Cloning, nucleotide sequence, and expression of the Pasteurella
RT   haemolytica A1 glycoprotease gene.";
RL   J. Bacteriol. 173:5597-5603(1991).
CC   -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF A MOIETY ON THE CORE
CC       OF THE LIPOPOLYSACCHARIDE MOLECULE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSEMBLE AT THE INNER SURFACE
CC       OF THE CYTOPLASMIC MEMBRANE (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U15958; AAA80283.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.
SQ   SEQUENCE   263 AA;  30871 MW;  B07C39EC1DA603A7 CRC64;

  Query Match             22.0%;  Score 321;  DB 1;  Length 263;
  Best Local Similarity   32.4%;  Pred. No. 1.9e-22;
  Matches   93;  Conservative   47;  Mismatches  115;  Indels   32;  Gaps    9;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           | |:|||| || ||| || | || : |||||||: |    |:| |  :    ::  |:
Db       1 MNNYVISLTSAQERRKHIEAEFGKQNIPFQFFDAITPDLIKEKAKAFNID--ISNTNLTK 58

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           | || :||  ||  |  : : || :||||: ||  | :       |:  : |::  :|:
Db      59 GEIACALSHIALWHLAKQQNLDYICIFEDDIYLGNNAFEL------LKTNYIPENTHIVK 112

Qy     121 LETMFMHVLTSPSGVADY--GGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERL 178
           |||:    :   :    |    | |  | | | |||||:: |   | ::   |
Db     113 LETLPFDRINRFNKTEKYILNRRLFK-LNSRHVGTAGYILTNKGAEFLINILKTLN---- 167

Qy     179 HPVDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDR--RLNR--- 233
           |:| ::|      :   | |::|||| |:       :  :   | :: ||  | |
Db     168 IPIDDLIFDEYLKIKEYKVLQMSPALCVQDF----ILNSKTNFKSSLQDDRAIRCNNEDK 223

Qy     234 -KQQWRDSPANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
           | | :    | |       |||| :|  :  :: :| :    : |:
Db     224 IKNQAKLKKPNYF-------LTKIQKELYRPFKQLKQKVFTTYITFK 263


RESULT 9
Y765_HAEIN
ID   Y765_HAEIN     STANDARD;      PRT;   282 AA.
AC   Q57125; O05033;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical glycosyl transferase HI0765 (EC 2.-.-.-).
GN   HI0765.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
```

```
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32760; AAC22423.1; -.
DR   TIGR; HI0765; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Hypothetical protein; Transferase; Glycosyltransferase;
KW   Complete proteome.
SQ   SEQUENCE   282 AA;  32481 MW;  D962E402AA207A18 CRC64;

  Query Match           18.5%;  Score 271;  DB 1;  Length 282;
  Best Local Similarity   29.0%;  Pred. No. 8.8e-18;
  Matches   80;  Conservative   43;  Mismatches  103;  Indels   50;  Gaps    9;

Qy       3 NHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHP--YLSG 60
           |:|||| :  :|| ||  .|| : |||:||||: | : :|    :    |   |   ||
Db      34 NYVISLTTEQKRRKHITEEFGKQNIPFEFFDAITP-DIIEETAKKFNITLDRSPKAKLSD 92

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
             |  |  :|| |||: ||:  : || :|||| |||: |:: |  |      :  |   |::
Db      93 GEIGCALSHIVLWDLALENNLNYINIFEDDIHLGENAKELLEID------YISDDIHVLK 146

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           ||         |  |       :|:   :  | ||| :: |  :: |:     | :
Db     147 LEANGKMFFKQPKSV-KCDRNVYPMTVKQ-SGCAGYTVTAKGAKYLLELVKNKPLD--VA 202

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQE--LHYAKFHDQNSALGSLIEHDRRLNRKQQWR 238
           || ::| :    :   : ||:| :| |:  ||
Db     203 VDSLVFEDFLHFKDYKIVQLSPGICVQDFVLH---------------------------- 234

Qy     239 DSPANTFKHRLIRALTKIGREREKRRQRREQLIGKI 274
             | | |:  |     : ||:|    ||:  :: ||
Db     235 --PDNPFESSL-----QEGRDRVHGNQRKSSILEKI 263


RESULT 10
LSG4_HAEIN
ID   LSG4_HAEIN     STANDARD;      PRT;   257 AA.
AC   P71398; Q48213;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   LSG locus putative protein 4.
GN   HI1697.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A2;
RA   McLaughlin R., Abu Kwaik Y., Young R., Spinola S., Apicella M.;
RT   "Characterization and sequence of the lsg locus from Haemophilus
RT   influenzae.";
RL   Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32842; AAC23343.1; -.
DR   EMBL; M94855; AAA24981.1; -.
DR   TIGR; HI1697; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Complete proteome.
FT   CONFLICT     17     17       E -> K (IN REF. 1).
FT   CONFLICT     30     30       V -> I (IN REF. 1).
FT   CONFLICT     46     46       A -> S (IN REF. 1).
FT   CONFLICT    102    102       S -> L (IN REF. 1).
FT   CONFLICT    127    127       K -> N (IN REF. 1).
FT   CONFLICT    130    130       D -> N (IN REF. 1).
FT   CONFLICT    134    134       F -> T (IN REF. 1).
FT   CONFLICT    153    153       D -> N (IN REF. 1).
FT   CONFLICT    214    214       I -> M (IN REF. 1).
SQ   SEQUENCE   257 AA;  29864 MW;  D1ECDE47763EE059 CRC64;

  Query Match             6.9%;  Score 100.5;  DB 1;  Length 257;
  Best Local Similarity   23.0%;  Pred. No. 0.046;
  Matches   65;  Conservative   42;  Mismatches  113;  Indels   63;  Gaps   15;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGI-PFQFFDAL--MPSERLERAMAELVPGLSAHPY 57
           :: ::|||    :||      |  :    || | |:  |  :  | |    :    || |
Db       2 LKKYLISLDKDIQRR---ELFFSQKNTEDFQVFSAINTMQKDWDELAAIFNIEQFKAH-Y 57

Qy      58 LSGVEK---ACFMSHAVLWEQALDEG----VPYIAVFEDDVLLGEGAEQFLAEDTWLQER 110
             | |    | :||  :::: :::       |  | ||| |     :| |     | |:
Db      58 FRNVTKGEIGCTLSHLSVYQKIVEDNDIAEDSYALVCEDDALFHSDFQQNLT--ALLSEK 115

Qy     111 FDP-------------DSAFVVRLETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGY 156
             :                |    :  | |   |   :|  :|   |||  :|   || ||
Db     116 LEAEIILLGQSKINDFNDFDLEINYPTTF-SFLCKKTGDVNY---AFP-YKSYFAGTVGY 170

Qy     157 IISRKAMRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHD 216
           :| : | | |: : :   |  |   | ::|                            |:
Db     171 LIKKSAARRFIQQISQNKPFWLAD-DFLLFEQ----------------------NFNI 205

Qy     217 QNSALGSL--IEHDRRLNRKQQWRDSPANTFKHRLIR-ALTKI 256
           :|  :  |  ||:   ::  :  | | :|    :|::  | ||
Db     206 RNKVVRPLIVIENPVLISNLESVRGSLSNNLLKKLMKYPLKKI 248


RESULT 11
EFG_CHLPN
ID   EFG_CHLPN      STANDARD;      PRT;   694 AA.
AC   Q9Z802; Q9JSE0;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Elongation factor G (EF-G).
```

```
GN   FUSA OR CPN0550 OR CP0202.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
CC       OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC       RIBOSOME.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       EF-G/EF-2 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001639; AAD18690.1; -.
DR   EMBL; AE002181; AAF38074.1; -.
DR   EMBL; AP002547; BAA98756.1; -.
DR   HSSP; P13551; 1DAR.
DR   PHCI-2DPAGE; Q9Z802; -.
DR   TIGR; CP0202; -.
DR   InterPro; IPR004540; EF-G.
DR   InterPro; IPR000640; EFG_C.
DR   InterPro; IPR004161; EFTU_D2.
DR   InterPro; IPR000795; EF_GTPbind.
DR   InterPro; IPR005225; Small_GTP.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF00679; EFG_C; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   TIGRFAMs; TIGR00484; EF-G; 1.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Elongation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   NP_BIND      18     25       GTP (BY SIMILARITY).
FT   NP_BIND      82     86       GTP (BY SIMILARITY).
FT   NP_BIND     136    139       GTP (BY SIMILARITY).
FT   CONFLICT    555    556       NT -> KK (IN REF. 3).
SQ   SEQUENCE   694 AA;  76587 MW;  6AD5941DBB52C87A CRC64;
```

```
  Query Match             6.0%;  Score 88;  DB 1;  Length 694;
  Best Local Similarity   23.5%;  Pred. No. 2.4;
  Matches   63;  Conservative   33;  Mismatches   78;  Indels   94;  Gaps   14;

Qy      41 LERAMAELVPGLSAHPYLSGVEKACFMSHAVLWEQALDEGVPYIA-VFEDDVLLGEGAEQ 99
           :||::  |   ::    :||||         :| ||   ||| || | : | :   ||:
Db      93 VERSLRVLDGAVAVFDAVSGVEP----QSETVWRQADKYGVPRIAFVNKMDRM---GADY 145

Qy     100 FLAEDTWLQERFDPDSAFVVRLETMFMHVLTSPSGVADYGGRAFPL-----LESEHCGTA 154
           | | :: ::|:                                 |  |||:      ||:  |
Db     146 FAAVES-MKEKL-------------------------GANAFPVHCPIGSESQFVGMV 177

Qy     155 GYIISRKAMRFFLD-----------------RFAVLPP---ERLHPVD-------LMMF 186
             :||:||: :|||                 | | |     | |  :|       : :
Db     178 D-LISQKAL-YFLDDTLGAKWEEKEISEDLKERCAELRANLLEELATIDESNEAFMMKVL 235

Qy     187 GNPDD----------REGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQ 236
           :||          |:|:  ::|| ||          :|  :  |      ||  :
Db     236 EDPDSITEDEIHQVMRKGVIENKINPVLCGTAF-------KNKGVQQL------LNVIVK 282

Qy     237 WRDSPANTFKHRLIRALTKIGREREKRR 264
           |  || :    | |   |       | ||
Db     283 WLPSPLDRGNIRGINLKTDQEISLEPRR 310
```

```
RESULT 12
DDL_SYNY3
ID   DDL_SYNY3      STANDARD;      PRT;   354 AA.
AC   P73632;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   D-alanine--D-alanine ligase (EC 6.3.2.4) (D-alanylalanine synthetase)
DE   (D-Ala-D-Ala ligase).
GN   DDL OR DDLA OR SLR1874.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA   Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA   Yamada M., Yasuda M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + D-alanine + D-alanine = ADP + phosphate
CC       + D-alanyl-D-alanine.
CC   -!- PATHWAY: ALONG WITH ALANINE RACEMASE, IT MAKES UP THE D-ALANINE
CC       BRANCH OF THE PEPTIDOGLYCAN BIOSYNTHETIC ROUTE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE D-ALANINE--D-ALANINE LIGASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90908; BAA17677.1; -.
DR   HSSP; P07862; 1IOV.
DR   InterPro; IPR000291; Dala_Dala_ligase.
DR   Pfam; PF01820; Dala_Dala_ligas; 1.
DR   TIGRFAMs; TIGR01205; D_ala_D_alaTIGR; 1.
DR   PROSITE; PS00843; DALA_DALA_LIGASE_1; 1.
```

```
DR   PROSITE; PS00844; DALA_DALA_LIGASE_2; 1.
KW   Ligase; Cell wall; Peptidoglycan synthesis; Complete proteome.
SQ   SEQUENCE   354 AA;  38774 MW;  871B82362A47315A CRC64;

  Query Match             5.9%;  Score 86.5;  DB 1;  Length 354;
  Best Local Similarity   24.3%;  Pred. No. 1.4;
  Matches   43;  Conservative   33;  Mismatches   60;  Indels   41;  Gaps   11;

Qy      74 EQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVRLETMFMHVLTSPS 133
           :| ||:|||:            | :   |   |   :| |::|   |:|  | :|  |:
Db      57 QQVLDQGVPW-----------GDQPVTAGQRW---QFPPEAA---RMEVWF-PILHGPN 97

Qy     134 GVADYGGRAFPLLESEHCGTAGYI-----ISRKAMRFFLDRFAVLPPERLHPVDL-MMFG 187
           |       | |::  : | :| :     : : ||:   :| | ||      |:  ::
Db      98 GEDGTVQGLFSLMQVPYVG-SGVLGSCVGMDKLAMKMVFER-AGLPQVNYMGVERGEIWS 155

Qy     188 NP----------DDREGMPVCQLNPALCAQELHYAKFHDQN---SALGSLIEHDRRL 231
           ||          : : | | | : ||     :  ||  :::   :|| :   :|||:
Db     156 NPCVFPALCEKIEAQVGYP-CFVKPANLGSSVGIAKVRNRSELEAALDNAASYDRRI 211


RESULT 13
PIFA_ECOLI
ID   PIFA_ECOLI     STANDARD;      PRT;   741 AA.
AC   P96329;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phage T7 exclusion protein.
GN   PIFA.
OS   Escherichia coli.
OG   Plasmid F.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / CR63;
RA   Shimizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.;
RT   "Complete nucleotide sequence of the F plasmid: its implications for
RT   organization and diversification of plasmid genomes.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 411-735 FROM N.A.
RC   STRAIN=K12;
RA   Molineux I.J.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   FUNCTION.
RX   MEDLINE=71240788; PubMed=5283385;
RA   Morrison T.G., Malamy M.H.;
RT   "T7 translational control mechanisms and their inhibition by F
RT   factors.";
RL   Nature New Biol. 231:37-41(1971).
CC   -!- FUNCTION: RESPONSIBLE FOR THE EXCLUSION OF PHAGE T7 BY PLASMID F.
CC       GROWTH OF BACTERIOPHAGE T7 IS INHIBITED IN CELLS OF E.COLI THAT
CC       CARRIES THE PLASMID F.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP001918; BAA97910.1; -.
DR   EMBL; U89941; AAB48033.1; -.
DR   EcoGene; EG40053; pifA.
KW   Plasmid; Complete proteome.
FT   CONFLICT    701    741       DQVSPFFDDWPAVEEKLKVMLSGNELTPEQEALKTALENDD
FT                                -> VTRSVRSLMTGQQSKRS (IN REF. 2).
SQ   SEQUENCE   741 AA;  84965 MW;  4C373DF929AFCD65 CRC64;

  Query Match             5.7%;  Score 83.5;  DB 1;  Length 741;
  Best Local Similarity   22.1%;  Pred. No. 6.9;
  Matches   75;  Conservative   30;  Mismatches  123;  Indels  111;  Gaps   16;

Qy      15 RAHIAATFGSRGIPF----QFFDALMPSE-----RLERAMAELVPGLSAHPYLSGVEKAC 65
           || ||    |  : |       | | ||:     || ||:|:|     |  :  :  |
Db     184 RAEIAGQLVSLDLKFIVVMDDLDRLEPSQVAEVFRLVRAVADL-------PRFTHI--LC 234

Qy      66 F----MSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQ------ERFDPDS 115
           :    ::|||  | ||:               |   ::| :   |       | ||  :
Db     235 YDRQIITHAV--EHALNI--------------EDGSRYLQKIIQLSFKLPRPEAFDLRN 277

Qy     116 AFVVRLETMFMHVLTSP--SG-------VADYGGRAFPLLESEHCGTAGYIISRKAMRFF 166
            |  | | ::  :   |  ||       | |  | |       |      |     || |
Db     278 EFRQRAEALYQQINNQPPDSGMVRDLIAVTDTYGAALSTPREIHQAINSLIFLYPGMRDF 337

Qy     167 LDRFAVLPPERLHPVDLMMFGNP----------------------------DDREGMPV 197
           |  |: |  : |:   ||                                 | |||:
Db     338 -----VYFPD-LCLLQLIRVTNPALYDWTEHYLTERSVIETGQGMLSDGEKADFREGLIR 391

Qy     198 CQLN----------------PALCAQELHYAK-FHDQNSALGSLIEHDRRLNRKQQWR-- 238
           |                  | :    |   |   :    :   :||:    ||
Db     392 CMKTFRASNADSFLTLADWIPGISGHNDEYLNLFEPVSEDFRHIQTTGKRLSSLTHWRYY 451

Qy     239 ---DSPANTFKHRLIRALTKIGREREKRRQRREQLIGKI 274
              || |       | | :     ||::|  | |: ||
Db     452 FAFSSPQNVLPPEFFRQLFEQAGVSEKQQQLSELLLSKI 490


RESULT 14
KBF1_MOUSE
ID   KBF1_MOUSE     STANDARD;      PRT;   971 AA.
AC   P25799;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nuclear factor NF-kappa-B p105 subunit (DNA-binding factor KBF1) (EBP-
DE   1) (NF-kappa-B1 p84/NF-kappa-B1 p98) [Contains: Nuclear factor NF-
DE   kappa-B p50 subunit].
GN   NFKB1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90367113; PubMed=2203532;
RA   Ghosh S., Gifford A.M., Riviere L.R., Tempst P., Nolan G.P.,
RA   Baltimore D.;
RT   "Cloning of the p50 DNA binding subunit of NF-kappa B: homology to
RT   rel and dorsal.";
RL   Cell 62:1019-1029(1990).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS P84 AND P98).
RX   MEDLINE=95059075; PubMed=7969179;
RA   Grumont R.J., Fecondo J., Gerondakis S.;
RT   "Alternate RNA splicing of murine nfkb1 generates a nuclear isoform
RT   of the p50 precursor NF-kappa B1 that can function as a
RT   transactivator of NF-kappa B-regulated transcription.";
RL   Mol. Cell. Biol. 14:8460-8470(1994).
RN   [3]
RP   SEQUENCE OF 365-971 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=94001687; PubMed=8398903;
RA   Gerondakis S., Morrice N., Richardson I.B., Wettenhall R., Fecondo J.,
RA   Grumont R.J.;
RT   "The activity of a 70 kilodalton I kappa B molecule identical to the
RT   carboxyl terminus of the p105 NF-kappa B precursor is modulated by
RT   protein kinase A.";
RL   Cell Growth Differ. 4:617-627(1993).
RN   [4]
RP   SEQUENCE OF 365-971 FROM N.A.
RX   MEDLINE=92191289; PubMed=1339305;
```

```
RA   Inoue J., Kerr L.D., Kakizuka A., Verma I.M.;
RT   "I kappa B gamma, a 70 kd protein identical to the C-terminal half of
RT   p110 NF-kappa B: a new member of the I kappa B family.";
RL   Cell 68:1109-1120(1992).
RN   [5]
RP   COTRANSLATIONAL FOLDING OF P105.
RX   PubMed=9529257;
RA   Lin L., DeMartino G.N., Greene W.C.;
RT   "Cotranslational biogenesis of NF-kappaB p50 by the 26S proteasome.";
RL   Cell 92:819-828(1998).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 39-364.
RX   MEDLINE=95132005; PubMed=7530332;
RA   Ghosh G., van Duyne G., Ghosh S., Sigler P.B.;
RT   "Structure of NF-kappa B p50 homodimer bound to a kappa B site.";
RL   Nature 373:303-310(1995).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 245-350.
RX   MEDLINE=98046092; PubMed=9384558;
RA   Huang D.B., Huxford T., Chen Y.Q., Ghosh G.;
RT   "The role of DNA in the mechanism of NFkappaB dimer formation:
RT   crystal structures of the dimerization domains of the p50 and p65
RT   subunits.";
RL   Structure 5:1427-1436(1997).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 245-363.
RX   MEDLINE=99081292; PubMed=9865694;
RA   Huxford T., Huang D.B., Malek S., Ghosh G.;
RT   "The crystal structure of the IkappaBalpha/NF-kappaB complex reveals
RT   mechanisms of NF-kappaB inactivation.";
RL   Cell 95:759-770(1998).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 39-350.
RX   MEDLINE=98111237; PubMed=9450761;
RA   Chen F.E., Huang D.B., Chen Y.Q., Ghosh G.;
RT   "Crystal structure of p50/p65 heterodimer of transcription factor NF-
RT   kappaB bound to DNA.";
RL   Nature 391:410-413(1998).
CC   -!- FUNCTION: Appears to have dual functions such as cytoplasmic
CC       retention of attached NF-kappaB proteins and generation of p50 by
CC       a cotranslational processing. The proteasome-mediated process
CC       ensures the production of both p50 and p105 and preserves their
CC       independent function, although processing of NFKB1/p105 also
CC       appears to occur posttranslationally. p50 binds to the kappa-B
CC       consensus sequence 5'-GGRNNYYCC-3', located in the enhancer region
CC       of genes involved in immune response and acute phase reactions.
CC       Play a role in the regulation of apoptosis.
CC   -!- FUNCTION: NF-kappa-B p98 subunit (but not p84 or p105) acts as a
CC       transactivator of NF-kappa B-regulated gene expression.
CC   -!- SUBUNIT: Active NF-kappa-B is a heterodimer of an about 50 kDa
CC       DNA-binding subunit and the weak DNA-binding subunit p65. Two
CC       heterodimers might form a labile tetramer. Also interacts with
CC       MAP3K8.
CC   -!- SUBCELLULAR LOCATION: Nuclear, but also found in the cytoplasm
CC       in an inactive form complexed to an inhibitor (I-kappa-B).
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; p105 (shown here), p84 and p98;
CC       are produced by alternative splicing.
CC   -!- INDUCTION: By phorbol ester and TNF-alpha.
CC   -!- DOMAIN: The C-terminus of p105 might be involved in cytoplasmic
CC       retention, inhibition of DNA-binding, and transcription
CC       activation.
CC   -!- DOMAIN: Glycine-rich region (GRR) appears to be a critical element
CC       in the generation of p50.
CC   -!- PTM: While translation occurs, the particular unfolded structure
CC       after the GRR repeat may promote the generation of p50 making it
CC       an acceptable substrate for the proteasome. The processed form is
CC       active and the unprocessed form acts as an inhibitor (I kappa B-
CC       like), being able to form cytosolic complexes with NF-kappa B,
CC       trapping it in the cytoplasm. Complete folding of the region
CC       downstream of the GRR repeat precludes processing.
CC   -!- PTM: Polyubiquitination seems to allow P105 processing.
CC   -!- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC   -!- SIMILARITY: CONTAINS 7 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M57999; AAA40415.1; -.
DR   EMBL; S66656; AAB28573.1; -.
DR   EMBL; S89033; AAB21851.1; -.
DR   PIR; A35697; A35697.
DR   PDB; 1NFK; 23-DEC-96.
DR   PDB; 1BFS; 28-JAN-98.
DR   PDB; 1BFT; 28-JAN-98.
DR   PDB; 1VKX; 13-JAN-99.
DR   PDB; 1IKN; 12-APR-99.
DR   TRANSFAC; T01924; -.
DR   MGD; MGI:97312; Nfkb1.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00531; death; 1.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   PRINTS; PR00057; NFKBTNSCPFCT.
DR   SMART; SM00248; ANK; 5.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 5.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS01204; REL_1; 1.
DR   PROSITE; PS50254; REL_2; 1.
KW   DNA-binding; Transcription regulation; Activator; Nuclear protein;
KW   Phosphorylation; Repeat; ANK repeat; Alternative splicing;
KW   3D-structure; Apoptosis.
FT   DOMAIN       40    365       REL-LIKE (RHD).
FT   REPEAT      538    567       ANK 1.
FT   REPEAT      577    606       ANK 2.
FT   REPEAT      610    639       ANK 3.
FT   REPEAT      646    675       ANK 4.
FT   REPEAT      680    710       ANK 5.
FT   REPEAT      714    743       ANK 6.
FT   REPEAT      767    797       ANK 7.
FT   DOMAIN      358    363       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      373    431       GLY-RICH.
FT   MOD_RES     335    335       PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT   MOD_RES     930    930       PHOSPHORYLATION (BY SIMILARITY).
FT   VARSPLIC    780    971       VFDILNGKPYEPVFTSDDILPQGDMKQLTEDTRLQLCKLLE
FT                                IPDPDKNWATLAQKLGLGILNNAFRLSPAPSKTLMDNYEVS
FT                                GGTIKELMEALQQMGYTEAIEVIQAAFRTPATTASSPVTTA
FT                                QVHCLPLSSSSTRQHIDELRDSDSVCDSGVETSFRKLSFTE
FT                                SLTGDSPLLSLNKMPHGYGQEGPIEGKI -> GT (IN
FT                                ISOFORM P84).
FT   VARSPLIC    860    971       VSGGTIKELMEALQQMGYTEAIEVIQAAFRTPATTASSPVT
FT                                TAQVHCLPLSSSSTRQHIDELRDSDSVCDSGVETSFRKLSF
FT                                TESLTGDSPLLSLNKMPHGYGQEGPIEGKI -> MNSGIVT
FT                                ASVTVVWRHPSANSALQSLLLETAHCYL (IN ISOFORM
FT                                P98).
SQ   SEQUENCE   971 AA;  105641 MW;  E03687260192A934 CRC64;

  Query Match             5.7%;  Score 83.5;  DB 1;  Length 971;
  Best Local Similarity   25.6%;  Pred. No. 9.8;
  Matches   41;  Conservative   19;  Mismatches   55;  Indels   45;  Gaps    7;

Qy       6 ISLASA--AERRAHIAATFGSRGIPFQFFDALMP----SERLERAMAELVPGLSAHPYLS 59
           ||||     |  ||: :|          :|   |    : |    :| |:    | | :
Db     694 ISLAGCLLLEGDAHVDST---------TYDGTTPLHIAAGRGSTRLAALLKAAGADPLVE 744

Qy      60 GVEKACFMSHAVLWEQA-LDEGV----------------------PYIAVFEDDVLLGEG 96
```

```
           |    :  :  ||:|  ||||                          ||  ||  | :| :|
Db    745 NFEPLYDLDDS--WEKAGEDEGVVPGTTPLDMAANWQVFDILNGKPYEPVFTSDDILPQG 802

Qy     97 AEQFLAEDTWLQ---:-ERFDPDSAFVVRLETMFMHVLTS 131
          : | ||| ||     |  |||  :'   : : : :| :
Db    803 DMKQLTEDTRLQLCKLLEIPDPDKNWATLAQKLGLGILNN 842


RESULT 15
RET_HUMAN
ID   RET_HUMAN      STANDARD;      PRT;  1114 AA.
AC   P07949;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Proto-oncogene tyrosine-protein kinase receptor ret precursor
DE   (EC 2.7.1.112) (C-ret).
GN   RET.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-280 FROM N.A.
RX   MEDLINE=89282215; PubMed=2660074;
RA   Takahashi M.;
RT   "Isolation of ret proto-oncogene cDNA with an amino-terminal signal
RT   sequence.";
RL   Oncogene 4:805-806(1989).
RN   [2]
RP   SEQUENCE OF 255-1114 FROM N.A.
RX   MEDLINE=90272230; PubMed=3078962;
RA   Takahashi M., Buma Y., Iwamoto T., Inaguma Y., Ikeda H., Hiai H.;
RT   "Cloning and expression of the ret proto-oncogene encoding a tyrosine
RT   kinase with two potential transmembrane domains.";
RL   Oncogene 3:571-578(1988).
RN   [3]
RP   SEQUENCE OF 588-1063 FROM N.A.
RX   MEDLINE=87257826; PubMed=3037315;
RA   Takahashi M., Cooper G.M.;
RT   "ret transforming gene encodes a fusion protein homologous to
RT   tyrosine kinases.";
RL   Mol. Cell. Biol. 7:1378-1385(1987).
RN   [4]
RP   REVIEW ON HSCR VARIANTS.
RX   MEDLINE=98023959; PubMed=9359036;
RA   Hofstra R.M.W., Osinga J., Buys C.H.C.M.;
RT   "Mutations in Hirschsprung disease: when does a mutation contribute to
RT   the phenotype.";
RL   Eur. J. Hum. Genet. 5:180-185(1997).
RN   [5]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97220587; PubMed=9067749;
RA   Eng C., Mulligan L.M.;
RT   "Mutations of the RET proto-oncogene in the multiple endocrine
RT   neoplasia type 2 syndromes, related sporadic tumours, and
RT   hirschsprung disease.";
RL   Hum. Mutat. 9:97-109(1997).
RN   [6]
RP   VARIANTS MEN2A/MTC W-611; S-618; R-620; Y-620 AND R-634.
RX   MEDLINE=93372843; PubMed=8103403;
RA   Donis-Keller H., Dou S., Chi D., Carlson K.M., Toshima K.,
RA   Lairmore T.C., Howe J.R., Moley J.F., Goodfellow P., Wells S.A. Jr.;
RT   "Mutations in the RET proto-oncogene are associated with MEN 2A and
RT   FMTC.";
RL   Hum. Mol. Genet. 2:851-856(1993).
RN   [7]
RP   VARIANTS G-618; 632-DVR-634; G-634; F-634; Y-634 AND S-634.
RX   MEDLINE=93275414; PubMed=8099202;
RA   Mulligan L.M., Kwok J.B.J., Healey C.S., Elsdon M.J., Eng C.,
RA   Gardner E., Love D.R., Mole S.E., Moore J.K., Papi L., Ponder M.A.,
RA   Telenius H., Tunnacliffe A., Ponder B.A.J.;
RT   "Germ-line mutations of the RET proto-oncogene in multiple endocrine
RT   neoplasia type 2A.";
RL   Nature 363:458-460(1993).
RN   [8]
RP   VARIANTS HSCR P-40; L-399; Q-762; P-765; Q-897; G-972 AND L-973.
RX   MEDLINE=95219414; PubMed=7704557;
RA   Yin L., Barone V., Seri M., Bolino A., Bocciardi R., Ceccherini I.,
RA   Pasini B., Tocco T., Lerone M., Cywes S., Moore S.,
RA   Vanderwinden J.-M., Abramowicz M.J., Kristoffersson U., Larsson L.T.,
RA   Hamel B.C.J., Silengo M., Martucciello G., Romeo G.;
RT   "Heterogeneity and low detection rate of RET mutations in Hirschsprung
RT   disease.";
RL   Eur. J. Hum. Genet. 2:272-280(1994).
RN   [9]
RP   VARIANT MEN2B THR-918.
RX   MEDLINE=94272459; PubMed=7911697;
RA   Eng C., Smith D.P., Mulligan L.M., Nagai M.A., Healey C.S.,
RA   Ponder M.A., Gardner E., Scheumann G.F., Jackson C.E., Tunnacliffe A.,
RA   Ponder B.A.J.;
RT   "Point mutation within the tyrosine kinase domain of the RET
RT   proto-oncogene in multiple endocrine neoplasia type 2B and related
RT   sporadic tumours.";
RL   Hum. Mol. Genet. 3:237-241(1994).
RN   [10]
RP   VARIANTS FMTC/MEN2A R-618; S-618; F-620; R-620; F-634; G-634 AND
RP   Y-634.
RX   MEDLINE=94348513; PubMed=7915165;
RA   Xue F., Yu H., Maurer L.H., Memoli V.A., Nutile-Mcmenemy N.,
RA   Schuster M.K., Browden D.W., Mao J.-I., Noll W.W.;
RT   "Germline RET mutations in MEN 2A and FMTC and their detection by
RT   simple DNA diagnostic tests.";
RL   Hum. Mol. Genet. 3:635-638(1994).
RN   [11]
RP   VARIANTS FMTC/MEN2A TYR-609; ARG-618; SER-618 AND SER-620.
RX   MEDLINE=95152521; PubMed=7849720;
RA   Blaugrund J.E., Johns M.M. Jr., Eby Y.J., Ball D.W., Baylin S.B.,
RA   Hruban R.H., Sidransky D.;
RT   "RET proto-oncogene mutations in inherited and sporadic medullary
RT   thyroid cancer.";
RL   Hum. Mol. Genet. 3:1895-1897(1994).
RN   [12]
RP   VARIANTS FMTC, AND VARIANTS MEN2A.
RX   MEDLINE=95179108; PubMed=7874109;
RA   Schuffenecker I., Billaud M., Calender A., Chambe B., Ginet N.,
RA   Calmettes C., Modigliani E., Lenoir G.M.;
RT   "RET proto-oncogene mutations in French MEN 2A and FMTC families.";
RL   Hum. Mol. Genet. 3:1939-1943(1994).
RN   [13]
RP   VARIANT HSCR W-609, VARIANT HSCR/MEN2A R-618, AND VARIANT HSCR/FMTC
RP   R-620.
RX   MEDLINE=95187155; PubMed=7881414;
RA   Mulligan L.M., Eng C., Attie T., Lyonnet S., Marsh D.J., Hyland V.J.,
RA   Robinson B.G., Frilling A., Verellen-Dumoulin C., Safar A.,
RA   Venter D.J., Munnich A., Ponder B.A.J.;
RT   "Diverse phenotypes associated with exon 10 mutations of the RET
RT   proto-oncogene.";
RL   Hum. Mol. Genet. 3:2163-2167(1994).
RN   [14]
RP   VARIANT MEN2B THR-918.
RX   MEDLINE=94159102; PubMed=7906866;
RA   Hofstra R.M.W., Landsvater R.M., Ceccherini I., Stulp R.P.,
RA   Stelwagen T., Luo Y., Pasini B., Hoeppener J.W.M., van Amstel H.K.P.,
RA   Romeo G., Lips C.J.M., Buys C.H.C.M.;
RT   "A mutation in the RET proto-oncogene associated with multiple
RT   endocrine neoplasia type 2B and sporadic medullary thyroid
RT   carcinoma.";
RL   Nature 367:375-376(1994).
RN   [15]
RP   VARIANTS HSCR PRO-765; GLN-897 AND GLY-972.
RX   MEDLINE=94159103; PubMed=8114938;
RA   Romeo G., Ronchetto P., Luo Y., Barone V., Seri M., Ceccherini I.,
RA   Pasini B., Bocciardi R., Lerone M., Kaarlainen H., Martucciello G.;
RT   "Point mutations affecting the tyrosine kinase domain of the RET
RT   proto-oncogene in Hirschsprung's disease.";
```

```
RL   Nature 367:377-378(1994).
RN   [16]
RP   VARIANTS HSCR LEU-32; LEU-64; GLN-330 AND LEU-393.
RX   MEDLINE=94159104; PubMed=8114939;
RA   Edery P., Lyonnet S., Mulligan L.M., Pelet A., Dow E., Abel L.,
RA   Holder S., Nihoul-Fkete C., Ponder B.A.J., Munnich A.;
RT   "Mutations of the RET proto-oncogene in Hirschsprung's disease.";
RL   Nature 367:378-380(1994).
RN   [17]
RP   VARIANT MEN2B THR-918.
RX   MEDLINE=94151373; PubMed=7906417;
RA   Carlson K.M., Dou S., Chi D., Scavarda N., Toshima K., Jackson C.E.,
RA   Wells S.A. Jr., Goodfellow P.J., Donis-Keller H.;
RT   "Single missense mutation in the tyrosine kinase catalytic domain of
RT   the RET protooncogene is associated with multiple endocrine neoplasia
RT   type 2B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:1579-1583(1994).
RN   [18]
RP   VARIANTS MTC; FMTC; MEN2A AND MEN2B.
RX   MEDLINE=96223053; PubMed=8625130;
RA   Komminoth P., Kunz E.K., Matias-Guiu X., Hiort O., Christiansen G.,
RA   Colomer A., Roth J., Heitz P.U.;
RT   "Analysis of RET protooncogene point mutations distinguishes heritable
RT   from nonheritable medullary thyroid carcinomas.";
RL   Cancer 76:479-489(1995).
RN   [19]
RP   VARIANTS MEN2A SER-618; SER-620; ARG-634 AND TYR-634.
RX   MEDLINE=95163936; PubMed=7860065;
RA   Takiguchi-Shirahama S., Koyama K., Miyauchi A., Wakasugi T., Oishi S.,
RA   Takami H., Hikiji K., Nakamura Y.;
RT   "Germline mutations of the RET proto-oncogene in eight Japanese
RT   patients with multiple endocrine neoplasia type 2A (MEN2A).";
RL   Hum. Genet. 95:187-190(1995).
RN   [20]
RP   VARIANTS HSCR L-20; S-93; Q-330; Y-609 AND R-620, AND VARIANT C-982.
RC   TISSUE=Blood;
RX   MEDLINE=95360000; PubMed=7633441;
RA   Angrist M., Bolk S., Thiel B., Puffenberger E.G., Hofstra R.M.W.,
RA   Buys C.H.C.M., Cass D.T., Chakravarti A.;
RT   "Mutation analysis of the RET receptor tyrosine kinase in Hirschsprung
RT   disease.";
RL   Hum. Mol. Genet. 4:821-830(1995).
RN   [21]
RP   VARIANTS HSCR.
RC   TISSUE=Leukocyte;
RX   MEDLINE=96090258; PubMed=7581377;
RA   Attie T., Pelet A., Edery P., Eng C., Mulligan L.M., Amiel J.,
RA   Boutrand L., Beldjord C., Nihoul-Fekete C., Munnich A., Ponder B.A.J.,
RA   Lyonnet S.;
RT   "Diversity of RET proto-oncogene mutations in familial and sporadic
RT   Hirschsprung disease.";
RL   Hum. Mol. Genet. 4:1381-1386(1995).
RN   [22]
RP   VARIANT MEN2B THR-918, AND VARIANT TYR-922.
RX   MEDLINE=96121602; PubMed=8595427;
RA   Kitamura Y., Scavarda N., Wells S.A. Jr., Jackson C.E.,
RA   Goodfellow P.J.;
RT   "Two maternally derived missense mutations in the tyrosine kinase
RT   domain of the RET protooncogene in a patient with de novo MEN 2B.";
RL   Hum. Mol. Genet. 4:1987-1988(1995).
RN   [23]
RP   VARIANT FMTC ASP-768.
RX   MEDLINE=95148221; PubMed=7845675;
RA   Eng C., Smith D.P., Mulligan L.M., Healey C.S., Zvelebil M.J.,

  Query Match             5.7%;  Score 83.5;  DB 1;  Length 1114;
  Best Local Similarity   19.1%;  Pred. No. 12;
  Matches   41;  Conservative   52;  Mismatches   67;  Indels   55;  Gaps   10;

Qy      62 EKACFMSHAVLWEQALDEGVPYIA----VFED----DVLLGEGAEQFLAEDTWLQERFDP 113
           |:|  |   : :   : :|: |:|    |  |    ::|: || :  :::    :: ::
Db     843 ERALTMGDLISFAWQISQGMQYLAEMKLVHRDLAARNILVAEGRKMKISDFGLSRDVYEE 902

Qy     114 DSAFVVR----------LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKA 162
           || :| |           :|::| |: |: | |  :|   : ::      | |
Db     903 DS-YVKRSQGRIPVKWMAIESLFDHIYTTQSDVWSFGVLLWEIV------TLGG------ 949

Qy     163 MRFFLDRFAVLPPERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELH------YAKFHD 216
                : :  :|||||   :|:  |:  :|         |  |::|::      : :  |
Db     950 -----NPYPGIPPERL--FNLLKTGHRMER---------PDNCSEEMYRLMLQCWKQEPD 993

Qy     217 QNSALGSLIEH-DRRLNRKQQWRDSPANTFKHRLI 250
           :      : :  :: : ::: : |  |:|    ||
Db     994 KRPVFADISKDLEKMMVKRRDYLDLAASTPSDSLI 1028


Search completed: May 27, 2003, 18:14:20
Job time : 14 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 27, 2003, 18:12:24 ; Search time 87 Seconds
(without alignments)
660.772 Million cell updates/sec

Title: US-10-007-267A-8
Perfect score: 1461
Sequence: 1 MQNHVISLASAAERRAHIAA.....REKRRQRREQLIGKIIVPFQ 279

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1408	96.4	279	2	Q93PR9	Q93pr9 neisseria g
2	1305	89.3	248	2	Q9RGN1	Q9rgn1 neisseria g
3	1284	87.9	275	2	Q9EVD4	Q9evd4 neisseria s
4	1241	84.9	275	2	Q93EK5	Q93ek5 neisseria m
5	1111.5	76.1	280	2	Q93PR7	Q93pr7 neisseria g
6	1040.5	71.2	280	2	Q9REX4	Q9rex4 neisseria g
7	1002.5	68.6	280	2	Q9RGM8	Q9rgm8 neisseria g
8	997.5	68.3	268	16	Q9JW65	Q9jw65 neisseria m
9	967.5	66.2	280	2	P96947	P96947 neisseria m
10	931.5	63.8	280	2	Q9EVD3	Q9evd3 neisseria s
11	912.5	62.5	280	2	Q93EK4	Q93ek4 neisseria m
12	471	32.2	266	2	Q8RKU0	Q8rku0 haemophilus
13	439	30.0	280	2	Q9L7A4	Q9l7a4 haemophilus
14	387	26.5	282	2	Q9ZFP8	Q9zfp8 haemophilus
15	349	23.9	208	16	Q9CLR8	Q9clr8 pasteurella
16	252	17.2	117	2	P96944	P96944 neisseria m
17	252	17.2	133	2	Q93EK8	Q93ek8 neisseria m
18	237.5	16.3	265	16	Q98C26	Q98c26 rhizobium l
19	231.5	15.8	268	16	Q8YIM5	Q8yim5 brucella me
20	140.5	9.6	242	2	Q8RJK5	Q8rjk5 vibrio chol
21	131.5	9.0	257	2	Q9ZIS2	Q9zis2 escherichia
22	129.5	8.9	560	4	Q9P226	Q9p226 homo sapien
23	128.5	8.8	517	4	Q9UMW5	Q9umw5 homo sapien
24	128	8.8	332	16	Q9ZLL7	Q9zll7 helicobacte
25	127	8.7	230	2	Q8RJL2	Q8rjl2 vibrio chol
26	121	8.3	1128	5	Q9VMQ6	Q9vmq6 drosophila
27	120	8.2	207	2	Q8RJM1	Q8rjm1 vibrio chol
28	116.5	8.0	721	16	Q8YG36	Q8yg36 brucella me
29	111.5	7.6	703	2	Q8VP06	Q8vp06 brucella ab
30	110.5	7.6	287	2	O06023	O06023 haemophilus
31	108.5	7.4	251	2	Q9ALY2	Q9aly2 campylobact
32	107	7.3	738	4	O60327	O60327 homo sapien
33	106.5	7.3	243	4	Q9H6F1	Q9h6f1 homo sapien
34	105	7.2	261	2	Q937Y0	Q937y0 edwardsiell
35	103	7.0	224	2	Q9X6Q5	Q9x6q5 actinobacil
36	102.5	7.0	231	4	Q8WUI9	Q8wui9 homo sapien
37	102.5	7.0	261	16	Q9CNC3	Q9cnc3 pasteurella
38	102	7.0	254	2	Q9ALZ3	Q9alz3 campylobact
39	99.5	6.8	273	16	O25500	O25500 helicobacte
40	99	6.8	492	5	Q18989	Q18989 caenorhabdi
41	98.5	6.7	224	2	Q47960	Q47960 haemophilus
42	98.5	6.7	273	2	Q9RHG8	Q9rhg8 helicobacte
43	97	6.6	476	11	Q9D3S3	Q9d3s3 mus musculu
44	96	6.6	247	2	Q48023	Q48023 haemophilus
45	96	6.6	744	17	Q8ZVG1	Q8zvg1 pyrobaculum

ALIGNMENTS

RESULT 1
Q93PR9
ID Q93PR9 PRELIMINARY; PRT; 279 AA.
AC Q93PR9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Glycosyl transferase LgtB.
GN LGTB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PID2;
RA Tong Y., Arking D., Reinhold V., Stein D.C.;
RT "Characterization of lipooligosaccharide structures found in Neisseria
RT gonorrhoeae PID2.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF313394; AAK70339.1; -.
DR InterPro; IPR002654; GT_25.
DR Pfam; PF01755; Glyco_transf_25; 1.
KW Transferase.
SQ SEQUENCE 279 AA; 31637 MW; 7297F13AFA732379 CRC64;

Query Match 96.4%; Score 1408; DB 2; Length 279;
Best Local Similarity 96.4%; Pred. No. 8.6e-121;
Matches 269; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           ||||||||||||||||||||||||||||||||||||||||:||||||||||||| ||:
```

```
Db    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISREAMRFFLDRFAVLPAERIKA 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          |||||||||||||||||||||||:| || ||||||||
Db    241 PANTFKHRLIRALTKIGREREKRRKRCEQTIGKIIVPFQ 279


RESULT 2
Q9RGN1
ID   Q9RGN1      PRELIMINARY;      PRT;   248 AA.
AC   Q9RGN1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Glycosyltransferase.
GN   LGTB.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1291;
RX   MEDLINE=20305049; PubMed=10844691;
RA   Harvey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
RA   Phillips N.J., Apicella M.A., Balke M.S.;
RT   "Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein
RT   receptor on human sperm.";
RL   Mol. Microbiol. 36:1059-1070(2000).
DR   EMBL; AF121135; AAF14360.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase.
SQ   SEQUENCE   248 AA;  28036 MW;  D9036FB07D8C9397 CRC64;

  Query Match             89.3%;  Score 1305;  DB 2;  Length 248;
  Best Local Similarity   99.6%;  Pred. No. 2e-111;
  Matches  247;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240

Qy    241 PANTFKHR 248
          ||||||||
Db    241 PANTFKHR 248


RESULT 3
Q9EVD4
ID   Q9EVD4      PRELIMINARY;      PRT;   275 AA.
AC   Q9EVD4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Galactosyl transferase.
GN   LGTB.
OS   Neisseria subflava.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=28449;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21142520; PubMed=11208792;
RA   Arking D., Tong Y., Stein D.C.;
RT   "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae.";
RL   J. Bacteriol. 183:934-941(2001).
DR   EMBL; AF240672; AAG09765.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase.
SQ   SEQUENCE   275 AA;  31441 MW;  D719F3815F64D14C CRC64;

  Query Match             87.9%;  Score 1284;  DB 2;  Length 275;
  Best Local Similarity   88.9%;  Pred. No. 1.9e-109;
  Matches  248;  Conservative    9;  Mismatches   18;  Indels    4;  Gaps    1;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||  ||| ||||||||||||||||: ||||||||||||||||
Db      1 MQNHVISLASAAERRAHITDTFGVRGIPFQFFDALMPSERLEQVMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:||||||:||| ||||||||||||:||||| ||||||||||||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYITVFEDDVLLGEGAEKFLAEDAWLQERFDPDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          ||||||||||||||||||||||| |||||||  |||||||:||:|||||||| |||| |||
Db    121 LETMFMHVLTSPSGVADYGGRALPLLESEHWGTAGYIISKKAIRFFLDRFAALPPEGLHP 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||  :  ||||:|||||||||||||||||| |||||||||||||| ||||||||||
Db    181 VDLMMFSDFFDREGVPVCQLNPALCAQELHYAKLHDQNSALGSLIEHDRLLNRKQQWRDS 240

Qy    241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
          ||||||| |||||||| ||||||||||||     |||||
Db    241 PANTFKRRLIRALTKISREREKRRQRREQ----FIVPFQ 275


RESULT 4
Q93EK5
ID   Q93EK5      PRELIMINARY;      PRT;   275 AA.
AC   Q93EK5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   LgtB.
GN   LGTB.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M978;
RX   MEDLINE=21467954; PubMed=11583844;
RA   Zhu P., Klutch M.J., Tsai C.-M.;
RT   "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
RT   Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RL   FEMS Microbiol. Lett. 203:173-177(2001).
DR   EMBL; AF355193; AAL12841.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
SQ   SEQUENCE   275 AA;  31577 MW;  A73037E0625EE7B3 CRC64;

  Query Match             84.9%;  Score 1241;  DB 2;  Length 275;
  Best Local Similarity   86.4%;  Pred. No. 1.6e-105;
  Matches  241;  Conservative   17;  Mismatches   17;  Indels    4;  Gaps    1;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||||||:||||||||||||||:||||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIAATFGARGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
```

```
          ||||||||||||:||||||:||||||||||||||||||:||||| ||:|||||||||:||
Db      61 VEKACFMSHAVLWKQALDEGLPYIAVFEDDVLLGEGAEKFLAEDAWLKERFDPDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||||||||||||| |||||||||| |||||||:||:||||  | :| ||:: |
Db     121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISKKAIRFFLQHFVMLQPEQIKP 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           :|||||  :  |:||||| |::|||||||||||||||||||||||||| |||||| |||
Db     181 IDLMMFSDFFDKEGMPVYQVSPALCAQELHYAKFHDQNSALGSLIEHDRLLNRKQQRRDS 240

Qy     241 PANTFKHRLIRALTKIGRERERKRRQRREQLIGKIIVPFQ 279
           |||||| |||||||| ||||||||||     |||||
Db     241 PANTFKRRLIRALTKISRERERKRRQRREQ----FIVPFQ 275


RESULT 5
Q93PR7
ID   Q93PR7      PRELIMINARY;      PRT;   280 AA.
AC   Q93PR7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Glycosyl transferase LgtE.
GN   LGTE.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PID2;
RA   Tong Y., Arking D., Reinhold V., Stein D.C.;
RT   "Characterization of lipooligosaccharide structures found in Neisseria
RT   gonorrhoeae PID2.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF313394; AAK70340.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase.
SQ   SEQUENCE   280 AA;  32047 MW;  4F8171AC1BACFB45 CRC64;

  Query Match           76.1%;  Score 1111.5;  DB 2;  Length 280;
  Best Local Similarity   76.7%;  Pred. No. 1.2e-93;
  Matches  224;  Conservative   19;  Mismatches   24;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           ||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||||||||||||||||||||||||||||||||||:|||||||||||||||| ||:
Db     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISREAMRFFLDRFAVLPAERIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||    |:||||| |::|||| ||||||||  ||| ||| :| ||   |
Db     181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLIQNSMLGSDLEKDREQGR------- 233

Qy     241 PANTFKHR--------LIRALTKIGRERERKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:: |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 6
Q9REX4
ID   Q9REX4      PRELIMINARY;      PRT;   280 AA.
AC   Q9REX4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   LgtE.
GN   LGTE.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FA19;
RA   Balthazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT   "Mutations in the lgt operon influence serum-resistance in
RT   gonococci.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF208063; AAF20992.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
SQ   SEQUENCE   280 AA;  32448 MW;  A20B34CD786942E3 CRC64;

  Query Match           71.2%;  Score 1040.5;  DB 2;  Length 280;
  Best Local Similarity   71.9%;  Pred. No. 3.7e-87;
  Matches  210;  Conservative   23;  Mismatches   34;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||||||||:|||||||||||||||||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWEQALDEGLPYIAVFEDDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:||||||||||||||||:||||||||||||||:
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPPERIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           ||||||    |:||||| |::|||| ||||||||  ||| ||| :| ||   |
Db     181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy     241 PANTFKHR--------LIRALTKIGRERERKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:: |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKVYGRRVILFK 280


RESULT 7
Q9RGM8
ID   Q9RGM8      PRELIMINARY;      PRT;   280 AA.
AC   Q9RGM8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Glycosyltransferase.
GN   LGTE.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1291;
RX   MEDLINE=20305049; PubMed=10844691;
RA   Harvey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
RA   Phillips N.J., Apicella M.A., Balke M.S.;
RT   "Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein
RT   receptor on human sperm.";
RL   Mol. Microbiol. 36:1059-1070(2000).
DR   EMBL; AF121135; AAF14363.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase.
SQ   SEQUENCE   280 AA;  32304 MW;  2EB5CDDE7D2164E6 CRC64;

  Query Match           68.6%;  Score 1002.5;  DB 2;  Length 280;
  Best Local Similarity   69.9%;  Pred. No. 1.1e-83;
  Matches  204;  Conservative   23;  Mismatches   40;  Indels   25;  Gaps    3;
```

```
Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:|||||||||:|||||||| |: ||:|||||||||:|||| |||||||
Db     61 VEKACFMSHAVLWKQALDEGVPYVAVFEDDVLFGKDAEKFLAEDTWLEERFDKDSAFVVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||  |:  |  | :|  |:|||||||||||||||||:|||||||||||| ||:
Db    121 LETMFAKVIVRPDKVLNYENRSFPLLESEHCGTAGYIISREAMRFFLDRFAVLPAERIKA 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          ||||||    |:||||| |::|||| |||||||| ||| ||| :| ||   |
Db    181 VDLMMFTYFFDKEGMPVYQVSPALCTQELHYAKFLSQNSMLGSDLEKDREQGR------- 233

Qy    241 PANTFKHR--------LIRALTKIGREREK-----RRQRREQLIGKIIVPFQ 279
             :||        | ||| | |||: |     |:   |:: |: :: |:
Db    234 -----RHRRSLKAMFDLKRALGKFGREKNKTMESQRQAELEKVYGRRVILFK 280


RESULT 8
Q9JW65
ID   Q9JW65      PRELIMINARY;      PRT;   268 AA.
AC   Q9JW65;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Lacto-N-neotetraose biosynthesis glycosyl tranferase (LgtH).
GN   LGTB2 OR NMA0527 OR LGTH.
OS   Neisseria meningitidis (serogroup A), and
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699, 487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=N.meningitidis (serogroup A);
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=N.meningitidis; STRAIN=A1;
RX   MEDLINE=21467954; PubMed=11583844;
RA   Zhu P., Klutch M.J., Tsai C.-M.;
RT   "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
RT   Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RL   FEMS Microbiol. Lett. 203:173-177(2001).
DR   EMBL; AL162753; CAB83819.1; -.
DR   EMBL; AF355194; AAL12844.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Complete proteome.
SQ   SEQUENCE   268 AA;  30444 MW;  8F43967D0530DD02 CRC64;

  Query Match             68.3%;  Score 997.5;  DB 16;  Length 268;
  Best Local Similarity   72.9%;  Pred. No. 3e-83;
  Matches  207;  Conservative   24;  Mismatches   32;  Indels   21;  Gaps    5;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||  ||| ||||||||||||||:||||||||||||||||||
Db      1 MQNHVISLASAAERRAHITDTFGVRGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:|||||:|||||||||||||||||||:||||| ||||||||||||:||
Db     61 VEKACFMSHAVLWKQALDEGVPYVAVFEDDVLLGKDAEKFLAEDTWLEERFDKDSAFIVR 120

Db     61 VEKACFMSHAVLWKQALDEGLPYIAVFEDDVLLGEGAEKFLAEDAWLQERFDPDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||||||||||||||| ||||||||||| |||:  |||||||||||| ||:||||||||||||||||
Db    121 LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMRFFLDRFAVLPSERLKA 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          || |:| :  |: || |||| ||||           |:  | | ::: |: : :   | |
Db    181 VDWMLFSSFLDKGGMTVCQLTPALCV----------QSETLPSQLKNGRQESYRN--RRS 228

Qy    241 PANTFKHRLIRALTKIGRE----REKRRQRR-EQLIGKIIVPFQ 279
          |    |  | ||| |||||    ||::||:: |: :|: :|||:
Db    229 P----KVLLKRALGKIGREIERARERKRQKKLEKHLGRHVVPFE 268


RESULT 9
P96947
ID   P96947      PRELIMINARY;      PRT;   280 AA.
AC   P96947;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Glycosyl transferase.
GN   LGTE.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=126E;
RX   MEDLINE=20055626; PubMed=10589709;
RA   Jennings M.P., Srikhanta Y.N., Moxon E.R., Kramer M., Poolman J.T.,
RA   Kuipers B., van der Ley P.;
RT   "The genetic basis of the phase variation repertoire of
RT   lipopolysaccharide immunotypes in Neisseria meningitidis.";
RL   Microbiology 145:3013-3021(1999).
DR   EMBL; U65788; AAB48387.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase.
SQ   SEQUENCE   280 AA;  32589 MW;  1ED681E62E08AFCA CRC64;

  Query Match             66.2%;  Score 967.5;  DB 2;  Length 280;
  Best Local Similarity   67.1%;  Pred. No. 1.8e-80;
  Matches  196;  Conservative   29;  Mismatches   42;  Indels   25;  Gaps    3;

Qy      1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
          |||||||||||||||||| |||:  |||||||||||| ||:||||||||||||||||
Db      1 MQNHVISLASAAERRAHIADTFGAHDIPFQFFDALMPSEALEQAMAELVPGLSAHPYLSG 60

Qy     61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
          |||||||||||||:||||||:||:|||||||||: ||:||||||||:||||:||:||
Db     61 VEKACFMSHAVLWKQALDEGLPYVAVFEDDVLLGKDAEKFLAEDTWLEERFDKDSAFIVR 120

Qy    121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
          |||||  |:  |  | :|  |:|||||||| ||||||||:|||||:|||| ||||:||
Db    121 LETMFAKVIVRPDKVLNYENRSFPLLESEHWGTAGYIISREAMRFFLERFAVLPAERIKA 180

Qy    181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
          || |||    |:|||||| |:|||||| ||||||||  :|| ||| :| ||   |
Db    181 VDWMMFTYFFDKEGMPVYQVNPALCTQELHYAKFLSKNSMLGSDLEKDREQER------- 233

Qy    241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
               :||        | ||| | |||::|| :|:     |:  |: :: |:
Db    234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKAYGRRVISFK 280


RESULT 10
Q9EVD3
ID   Q9EVD3      PRELIMINARY;      PRT;   280 AA.
AC   Q9EVD3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
```

```
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Galactosyl transferase.
GN   LGTE.
OS   Neisseria subflava.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=28449;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21142520; PubMed=11208792;
RA   Arking D., Tong Y., Stein D.C.;
RT   "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae.";
RL   J. Bacteriol. 183:934-941(2001).
DR   EMBL; AF240672; AAG09766.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase.
SQ   SEQUENCE   280 AA;  32762 MW;  85CF77E94F215842 CRC64;

  Query Match             63.8%;  Score 931.5;  DB 2;  Length 280;
  Best Local Similarity   65.4%;  Pred. No. 3.5e-77;
  Matches  191;  Conservative   29;  Mismatches   47;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||||||||:|||||||||||||||| | | ||||||||:    ||
Db       1 MQNHVISLASAAERRAHIAATFGARGIPFQFFDALMPSEELNRMMAELVPGLAKQHLLSE 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||:||||||:||:||||||||||: ||:||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWKQALDEGLPYVAVFEDDVLLGKDAEKFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:||||||| |||||||||:||||||:|||||| | :
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHWGTAGYIISREAMRFFLERFAVLPAEWIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           || |||    |:||||| |:||||| |||||||| :|| ||| :| ||   |
Db     181 VDWMMFTYFFDKEGMPVYQVNPALCTQELHYAKFLSKNSMLGSDLEKDREQER------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:  |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKAYGRRVISFK 280


RESULT 11
Q93EK4
ID   Q93EK4      PRELIMINARY;      PRT;   280 AA.
AC   Q93EK4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   LgtE.
GN   LGTE.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M978;
RX   MEDLINE=21467954; PubMed=11583844;
RA   Zhu P., Klutch M.J., Tsai C.-M.;
RT   "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
RT   Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RL   FEMS Microbiol. Lett. 203:173-177(2001).
DR   EMBL; AF355193; AAL12842.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
SQ   SEQUENCE   280 AA;  32898 MW;  504256FF55D92F97 CRC64;

  Query Match             62.5%;  Score 912.5;  DB 2;  Length 280;
  Best Local Similarity   64.4%;  Pred. No. 1.9e-75;
  Matches  188;  Conservative   28;  Mismatches   51;  Indels   25;  Gaps    3;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
           |||||||||||||||||||||||:|||||||||||||||| | | ||||||||:    ||
Db       1 MQNHVISLASAAERRAHIADTFGRHDIPFQFFDALMPSEELNRVMAELVPGLAKQHLLSE 60

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
           ||||||||||||:||||||:||:||||||||||: ||:||||||||:|||| ||||:||
Db      61 VEKACFMSHAVLWKQALDEGLPYVAVFEDDVLLGKDAEKFLAEDTWLEERFDKDSAFIVR 120

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP 180
           |||||  |:  |  | :|  |:||||||| |||||||||:||||||:|||||| | :
Db     121 LETMFAKVIVRPDKVLNYENRSFPLLESEHWGTAGYIISREAMRFFLERFAVLPAEWIKA 180

Qy     181 VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS 240
           || |||    |:||||| |:||||| |||||||| :|| ||| :| ||   |
Db     181 VDWMMFTYFFDKEGMPVYQVNPALCTQELHYAKFLSKNSMLGSDLEKDREQER------- 233

Qy     241 PANTFKHR--------LIRALTKIGREREKRRQRR-----EQLIGKIIVPFQ 279
                :||        | ||| | |||::|| :|:     |:  |: :: |:
Db     234 -----RHRRSLKVMFDLKRALGKFGREKKKRMERQRQAELEKAYGRRVISFK 280


RESULT 12
Q8RKU0
ID   Q8RKU0      PRELIMINARY;      PRT;   266 AA.
AC   Q8RKU0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative galactosyl transferase Lic2B.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=G622;
RA   Pettigrew M.M., Foxman B., Marrs C.F., Gilsdorf J.R.;
RT   "Identification of LOS biosynthesis gene lic2B as a putative virulence
RT   factor in strains of non-typeable Haemophilus influenzae that cause
RT   otitis medi.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY091470; AAM12037.1; -.
KW   Transferase.
SQ   SEQUENCE   266 AA;  31182 MW;  B21CFA372200E2DC CRC64;

  Query Match             32.2%;  Score 471;  DB 2;  Length 266;
  Best Local Similarity   39.8%;  Pred. No. 4.4e-35;
  Matches  111;  Conservative   46;  Mismatches  104;  Indels   18;  Gaps    6;

Qy       3 NHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSGVE 62
           |:|||| || :|| ||   |  : |||:||||| ||: |   : : :| |  |  |:  |
Db       4 NYVISLTSAYQRREHIQKEFSQQNIPFEFFDALKPSKELTSLIEKFIPNL-LHAKLTEGE 62

Qy      63 KACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVRLE 122
           ||||||| :||:: | | :||| :||||||||| |::||||| ||:|||     |::| |
Db      63 KACFMSHYMLWQKCLSEDLPYIYIFEDDVLLGENADKFLAEDEWLEERFKQTDKFILRFE 122

Qy     123 TMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHPVD 182
           |            :  | ||    | ||:|| |||:|||:|::        |    |  :|
Db     123 TFLNFSKCKDKKIKPYSGRKILKLVSENCGAAGYVISREAVKQLSAHICSLTSNHLLAID 182

Qy     183 LMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRR--LNRKQQWRDS 240
           |:||    :       |::| :| ||    : : ::  | | :| :|:  |: |::
Db     183 LLMF----NIFNQSTYQVSPGVCVQE---GQLYPKDIKLHSQLETERQKYLSVKKK---- 231

Qy     241 PANTFKHRLIRALTKIGREREKRRQRREQLIGKIIVPFQ 279
             | |  ||    |   ::  |:  |:  | | ||||:
Db     232 --RTLKTVLISLAGK--PKKILRKIYRKLFISKHIVPFR 266


RESULT 13
Q9L7A4
ID   Q9L7A4      PRELIMINARY;      PRT;   280 AA.
AC   Q9L7A4;
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Galactosyltransferase II.
GN   LGTB.
OS   Haemophilus ducreyi.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=730;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35000HP;
RX   MEDLINE=20200369; PubMed=10735874;
RA   Sun S., Schilling B., Tarantino L., Tullius M.V., Gibson B.W.,
RA   Munson R.S. Jr.;
RT   "Cloning and characterization of the lipooligosaccharide
RT   galactosyltransferase II gene of Haemophilus ducreyi.";
RL   J. Bacteriol. 182:2292-2298(2000).
DR   EMBL; AF224466; AAF32397.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase.
SQ   SEQUENCE   280 AA;  33369 MW;  0639F668CF1F5E6E CRC64;

  Query Match             30.0%;  Score 439;  DB 2;  Length 280;
  Best Local Similarity   35.7%;  Pred. No. 4e-32;
  Matches  100;  Conservative   59;  Mismatches  111;  Indels   10;  Gaps    4;

Qy       3 NHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSGVE 62
Db       8 NYVISLKTADARRQHIIQEFAKHNIPFQFFDACSIENNLYMDIEKHLPML-LNSKLSNSE 66

Qy      63 KACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVRLE 122
Db      67 KGCLMSHFLLWKKCVLDDIPYMTIFEDDIILSDESNEFISDYSWVNNRFYEQKEILIKFE 126

Qy     123 TMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHPVD 182
Db     127 TFLMPVIRKFTKINEYKERAFNQLISRHFGTASYLISKEAAKYLILLYEKLPADELIAVD 186

Qy     183 LMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDSPA 242
Db     187 ESIFNVLLNDKNLCVYQLEPAICVQEL---QFNKENSLLISQIEEDRYKNQKQ---NKPR 240

Qy     243 NTFKHRLIRA---LTKIGREREKRRQRREQLIGKIIVPFQ 279
Db     241 KTLKQRIIKVCYNINRIFKKIKKTKLSDEKDINSDIVKFK 280


RESULT 14
Q9ZFP8
ID   Q9ZFP8      PRELIMINARY;      PRT;   282 AA.
AC   Q9ZFP8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Glycosyltransferase homolog Lob2A.
GN   LOB2A.
OS   Haemophilus somnus.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=731;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=738;
RX   MEDLINE=20072700; PubMed=10603403;
RA   Wu Y., McQuiston J.H., Cox A., Pack T.D., Inzana T.J.;
RT   "Molecular cloning and mutagenesis of a DNA locus involved in
RT   lipooligosaccharide biosynthesis in Haemophilus somnus.";
RL   Infect. Immun. 68:310-319(2000).
DR   EMBL; AF096997; AAC83803.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Transferase.
SQ   SEQUENCE   282 AA;  33687 MW;  494094032C7024DD CRC64;

  Query Match             26.5%;  Score 387;  DB 2;  Length 282;
  Best Local Similarity   35.5%;  Pred. No. 2.3e-27;
  Matches  102;  Conservative   47;  Mismatches  120;  Indels   18;  Gaps    6;

Qy       1 MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSG 60
Db       1 MFNYVISLKSAVDRRQHIEKEFSQQHIPYTFFDAFQLNEENKHILQHYVPALS-KSNLAN 59

Qy      61 VEKACFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR 120
Db      60 KLKSCFMSHVLLWQKCVDENLPFIAIFEDDIFLSSVANKFLSEYDWLKERLAFNQYFILR 119

Qy     121 LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFL------DRFAVLP 174
Db     120 LETIYMKSEYCKTKIPDYQNRKFNRLMSGHFGMGGYILSQVAAKKLLALFKNTDIF----- 175

Qy     175 PERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRK 234
Db     176 ---IEEIDQIIFKQLLKDEQYPVYQCEPAIIIQDCILNK--DNCTLTSSLAPERERERER 230

Qy     235 QQWRDSPAN-TFKHRLIRALTK-IGREREKRRQRREQLIGKIIVPFQ 279
Db     231 ERERESKNNKKAKKNLWRSISKEIGRFRRRIFERLSSHKILIIISLQ 277


RESULT 15
Q9CLR8
ID   Q9CLR8      PRELIMINARY;      PRT;   208 AA.
AC   Q9CLR8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein PM1141.
GN   PM1141.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR   EMBL; AE006155; AAK03225.1; -.
DR   InterPro; IPR002654; GT_25.
DR   Pfam; PF01755; Glyco_transf_25; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   208 AA;  24649 MW;  29625BA115E4047A CRC64;

  Query Match             23.9%;  Score 349;  DB 16;  Length 208;
  Best Local Similarity   37.3%;  Pred. No. 4.7e-24;
  Matches   82;  Conservative   40;  Mismatches   74;  Indels   24;  Gaps    6;

Qy      72 LWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVRLETMFMHVLTS 131
Db       1 MWHKCIQDNLPYIAIFEDDILLGRDARTFLAEDEWLFSRFNCDDIFIIRLETFLQETICE 60

Qy     132 --PSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHPVDLMMFGNP 189
Db      61 ALPNPVS-YCGRDFLALKDEHLGTAGYIISLGAAKYLLEIFKNMESNNIFPIDHLIFNRF 119

Qy     190 DDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDR-------RLNRKQQWRDSPA 242
Db     120 LAGEELMVYQLSPALCIQEL---QLNENESLLDSQLESERKNYRLAEKARKKKTWREKVY 176

Qy     243 NTFKHRLIRALTKIGREREKRRQRRE---QLIGKIIVPFQ 279
Db     177 HIF--------TKPQRMLKKRKERAEKNAKMKLKCIVKFE 208
```

Search completed: May 27, 2003, 18:17:18
Job time : 90 secs